BASIC PRINCIPLES OF ELECTRONICS

VOLUME 2 SEMICONDUCTORS

J. JENKINS AND W.H. JARVIS

The Commonwealth and International Library of Science Technology Engineering and Liberal Studies

JENKINS/JARVIS BASIC PRINCIPLES OF ELECTRONICS VOLUME 2 SEMICONDUCTORS

THE COMMONWEALTH AND INTERNATIONAL LIBRARY

Joint Chairman of the Honorary Editorial Advisory Board

PHYSICS DIVISION

General Editor: W. ASHHURST

BASIC PRINCIPLES OF ELECTRONICS

VOLUME 2

SEMICONDUCTORS

BASIC PRINCIPLES OF ELECTRONICS

VOLUME 2 SEMICONDUCTORS

J. Jenkins, B.Sc., A.Inst.P.
Lecturer in Electrical Engineering
South Devon Technical College

W. H. Jarvis, M.A. (Oxon), A.Inst.P.
Head of Physics, Rannoch School, Perthshire

PERGAMON PRESS
OXFORD · NEW YORK · TORONTO
SYDNEY · BRAUNSCHWEIG

Pergamon Press Ltd., Headington Hill Hall, Oxford
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523
Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1
Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

First edition 1971

Library of Congress Catalog Card No. 66–23849

Printed in Hungary

08 016119 7 (flexicover)
08 016118 9 (hard cover)

To L. H. J.

And A. M. J.

CONTENTS

(E denotes relevant experiment at end of Chapter)

PREFACE

THIS second volume, which is confined to semiconductors, is intended as an introduction to the subject for first degree and technical college students, and specialist sixth formers. No knowledge of calculus is needed, so it will also be of value to the non-specialist who wants to know how these devices work. Some knowledge of electricity and thermionic valve techniques (Vol. 1) is assumed.

We have endeavoured to use SI (Système International) units, symbols, and abbreviations throughout.*

The material in this volume has been the basis of short courses given to a wide range of technical college students; and the manuscript has been read, and much useful criticism offered, by two (at the time) sixth-form Physics pupils, David Barton and Huw Lloyd. Their help is greatly appreciated.

The authors wish to acknowledge the help given in the form of apparatus lent or supplied, and up-to-date advice given, by the manufacturers and suppliers whose addresses appear below.

The authors are grateful for permission to reprint questions from examinations set by the following authorities: The Institute

* The international symbol for a resistor is introduced in figures II. 27, II. 31 and II. 32.

of Physics and Physical Society (Grad. Inst. P), the Union of Lancashire and Cheshire Institutes (U.L.C.I.), the Council of Engineering Institutions (C. of E.I.), the University of London (U.L.), the City and Guilds of London Institute (C.G.L.I.) and the Institution of Electrical Engineers (I.E.E.).

We also wish to thank R. G. Hibberd for permission to use several of his diagrams of Texas instruments devices.

Finally, we apologize for any errors which have escaped correction, and would be very grateful if readers would bring them to our notice.

J. J.
W. H. J.

Torquay,

Mullard Ltd., Mullard House, Torrington Place, London W.C.1.

Texas Instruments Ltd., Manton Lane, Bedford.

Rainbow Radio ("Unilab") Ltd., Mincing Lane, Blackburn, Lancs.

Proops Brothers Ltd., 52 Tottenham Court Road, London W.1 P OBA.

W. B. Nicolson (Scientific Instruments) Ltd., Industrial Estate, Thornliebank, Glasgow.

Radionics Ltd., St. Lawrence House, Broad St., Bristol BS1 2HF.

CHAPTER 1

CONDUCTION IN THE SOLID STATE

1.1. HISTORICAL

The *Philosophical Transactions of the Royal Society* in the early part of the eighteenth century carry one of the first references to experiments on the electrically conductive properties of various substances. About 1730 Stephen Gray, whilst researching at Charterhouse School, then in London, had found that electric charges flowed with far less difficulty through a brass wire than through a silk thread.

Over 200 years later we do not find this so surprising. Both the brass and the silk consist of atoms, and, as the atoms consist of charged particles, it only remains to set these charges in motion and an electric current will result. We need to be able to describe fairly accurately the circumstances in which these charges can flow through matter, and so explain the existence of conductors, semiconductors and insulators.

It is very profitable to reconsider Faraday's experimental work on the electrolysis of solutions of salt. Arrhenius concluded that common salt (for example) when in solution splits up into positive sodium ions and negative chlorine ions, and these ions swim about independently. If a pair of electrodes be dipped

into the solution (see Fig. 1.1), the electric field created will cause the positive ions to drift towards the cathode (−), and the negative ions to drift towards the anode (+). So *two* currents act, but the negative ion current drifting to the anode can be thought

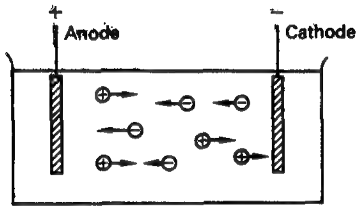


FIG. 1.1. Positive and negative charge carriers in electrolysis

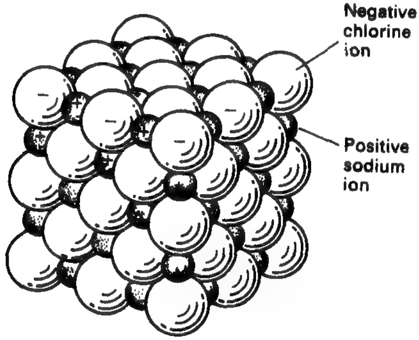


FIG. 1.2. A sodium chloride crystal (after Barlow). (Courtesy of Alan Holden)

of as a *second* positive ion current going to the cathode; and the total current is the sum of the currents carried by the two different ions.

About 1890 Barlow suggested that a crystal of solid sodium chloride is made up of ions closely packed together. More recent work with X-ray diffraction has shown that he was quite right; that it is the molecules, not the structure, which must be discarded. Figure 1.2 represents Barlow's picture of a sodium chloride

crystal, and it might seem surprising at first glance that the crystal is a good insulator. But the ions present have very little room to move and consequently very little current can flow. We picture the ions vibrating about their equilibrium positions until the temperature is raised sufficiently for the amplitude of the vibrations to knock the ions so hard that they can pass by one another. Instantly the crystal melts and conducts electricity.

We have so far presented a valid picture explaining why an ionic crystal should be a perfect insulator. In practice crystals *do* conduct slightly, and we now have to explain how it is possible for them to do so.

1.2. "VACANCIES" IN CRYSTALS

A crystal begins to form when a small number of atoms happen to arrange themselves in the right order in a suitable solution or molten material: then, if the existing conditions are just right, other atoms meeting the minute crystal will attach themselves to it and the crystal "grows", propagating outwards in all directions with the same geometrical pattern. It is unlikely that one could witness the actual rate of growth, which is typically about one-millionth of an inch in a second. On the microscopic scale, however, this is very fast and represents perhaps 100 new layers of atoms; and it is not surprising if there is an occasional "vacancy" where there should be an atom. In a perfectly ordered crystal with no vacancies it is most unlikely that atoms would have sufficient kinetic energy at normal temperatures to "squeeze" past one another; but in any practical crystal with "vacancies" in the structure it is quite easy for an atom next to a vacancy to slip into it, leaving a new vacancy behind; and as the atom moves one way the vacancy moves the opposite way. Clearly the

vacancy behaves as if it has a charge equal and opposite to the missing ion; and one can think of the vacancy as behaving like a charged particle. Figure 1.3 shows that one can think of conduction in a crystal as the successive movements through short distances to the *right* of many + ions, or as the continuous movement through a long distance to the *left* of one − ion.

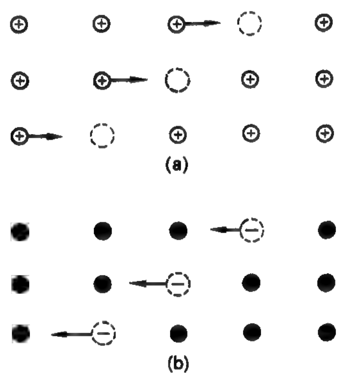


FIG. 1.3. Electric conduction in a crystal

Both interpretations are equivalent mathematically, but often the latter is more convenient. Of course the electrical conductivity so produced is very small indeed because the charge carriers involved in a crystal (the ions) are small in number and quite massive; but a minute current can flow through an ionic crystal in this way. In a metal, however, the density of charge carriers (electrons) is far greater, and electrons are far less massive than the ions, so it is not surprising that the electrical conductivity can be of the order of 10^{24} times greater in a metal than in an ionic crystal.

1.3. DRUDE'S THEORY OF METALLIC CONDUCTION

As far back as 1756 Franklin had suggested that "electric matter" consisted of minute particles, and in 1875 Weber introduced the concept of *charge* to this picture. When in 1897, Thomson identified the electron as a charged particle, many attempts were made to explain electrical conductivity in terms of these electrons. Drude's theory of 1900 was the most successful, and remains useful to this day, although superseded by more recent quantum mechanical theories. He appreciated that in order to explain the conductivity of a metal it was insufficient to consider the *number* of charged carriers alone; the ease with which they flowed was important, too. He adopted the basic idea of the kinetic theory of gases, which was pioneered by Maxwell and Boltzmann a half century before. Drude pictured the ions left behind by electrons as obstacles in a dense gas, the "atoms" of which were electrons; and as the electrons moved through the gas they sometimes would lose kinetic energy when they hit an ion, but at other times they might gain kinetic energy from the vibrations of the obstacle. In either case the electrons would be moving quite fast but irregularly, as Figure 1.4(a) shows. The picture is different when an electric field is applied, because a charged particle experiences a force in the direction of the electric field, and so, superimposed on the fast irregular motion of before, is a small "drift" velocity in one direction. This constitutes the current which the applied field has caused (see Fig. 1.4(b)).

In 1827 G. S. Ohm published a paper describing the results of many years work in electricity. He had found that the current flowing in a connection was directly proportional to the potential difference across the ends of the connection, at a certain

temperature. Since then experiments have been conducted for a very wide variety of materials and with a great range of p.d.s, and, provided the connection does not contain junctions between different metals, Ohm's law appears to be obeyed quite generally. It is particularly reliable for metallic conductors, deviations from the rule only approaching 1% in silver at current densities greater than 10^{10}A m^{-2}.

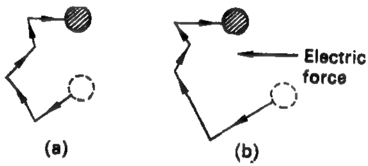


FIG. 1.4. Random thermal motion of an electron: (a) without an applied field, (b) with an applied field

The reader will be familiar with Ohm's law in the form of $I = V/R$, where I = current, V = voltage, and R = resistance of wire of length l, cross-sectional area A and resistivity ϱ (rho). By the definition of resistivity

$$R = \varrho \cdot \frac{l}{A}$$

So
$$I = \frac{V}{R} = \frac{1}{\varrho} \cdot \frac{V}{l} \cdot A$$

and
$$\frac{I}{A} = \frac{1}{\varrho} \times \frac{V}{l}$$

But I/A = current per unit area
= current density J
and $1/\varrho$ = conductivity σ (sigma)
and V/l = voltage gradient
= electric field intensity E.

So we write Ohm's law in the form

$$J = \sigma E \tag{1.1}$$

as this will enable us to arrive at an atomic interpretation of it. In the case of single crystals of a given substance σ varies and is dependent on the direction in which it is measured. For polycrystalline materials, however, eqn. (1.1) is valid, because then the value of σ is constant for a given material and is inde-

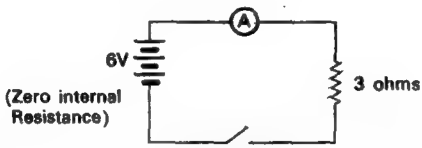


FIG. 1.5. The paradox: when the switch is closed the current of 2 A does not increase continuously

pendent of the direction in which it is measured. We shall assume that the materials considered in this book are isotropic, and that eqn. (1.1) holds. The reader would be correct in assuming that when an electric field is applied to a specimen there is a moment's delay before the current density reaches its equilibrium value, but as this time interval is about 10^{-14} sec it is reasonable to neglect it; there is a far more important point, however, which must be considered.

It is an easily verified experimental fact that if a steady direct p.d. is applied to a circuit, a current of constant magnitude will flow, e.g. in the circuit of Figure 1.5, a steady current of 2 amps will be indicated on the ammeter when the switch is closed. We certainly do not find the current increasing continuously! Yet this is just what eqn. (1.1) predicts. Let us see how this paradox arises.

1.4. THE PARADOX

When an electric field of strength E_x is applied in the x direction, an electron of charge $-e$ in this field will experience a force F_x given by

$$F_x = -eE_x \tag{1.2}$$

If the mass of the electron is m, the acceleration $(a_x)_{\text{field}}$ it experiences is:

$$(a_x)_{\text{field}} = -\frac{e}{m}E_x \tag{1.3}$$

from Newton's second law of motion, and the subscript "field" indicates that the acceleration is the rate of change of velocity u_x with time due to the interactions of the *electrons with the field only.*

If, in the electron gas, there are n free electrons per cubic metre, each of charge $-e$ coulombs, the total charge carried per unit volume is $-ne$ coulombs m^{-3}. But the number of electrons crossing unit area per second is equal to the number of electrons contained in a volume $u_x\,\text{m}^3$. So we can write

$$J = -neu_x \tag{1.4}$$

But we know from eqn. (1.3) that u_x is increasing, so J must be increasing, which is *not* observed experimentally. We are forced to conclude that there is another process at work here, and experiment requires that

$$(a_x)_{\text{field}} + (a_x)_{\text{other}} = 0 \tag{1.5}$$

To suggest that this "other" acceleration arises from interactions of electrons with one another is useless because such colli-

sions conserve momentum and would not alter u_x. However, we can find some clues because in the *vacuum* of, say, a diode valve the electrons will accelerate under the action of a constant applied p.d., in accordance with eqn. (1.3); and we know, too, that the passage of an electric current through a substance (as opposed to a vacuum) liberates heat in that substance, and this increase in vibrational kinetic energy of the molecules can only come from the collisions of passing electrons with the molecules, because the molecules could not gain energy directly from the applied field. So we conclude that the "other" process referred to in eqn. (1.5) is an "electron–lattice" interaction, that the electrons in the electron gas are not entirely free but must interact with the ions in the crystal lattice.

1.5. MEAN FREE TIME

We have seen that we must suppose that every now and then an electron has an *effective collision* with an ion, otherwise its velocity would increase continuously and so would the current. In the instant of time just after an effective collision the velocity of the electron is zero, and the electron has lost all "memory" of what the forces did to it in the past. It now begins to accelerate under the action of the applied field, and continues to do so until the next effective collision. Suppose that its velocity *just* before this collision is u_F, i.e. its final velocity. Clearly its average or drift velocity u_D is given by

$$u_D = \frac{u_F}{2} \tag{1.6}$$

We cannot substitute this result directly into eqn. (1.3) as that is an equation in acceleration, not velocity. However, we can

relate these two concepts by introducing the concept of *mean free time*, t. The times taken by various electrons to accelerate from zero velocity to u_F will obviously vary considerably; by mean free time we mean the average such time for all electrons.

So
$$(a_x)_{\text{field}} = \frac{u_F}{t} \tag{1.7}$$

Substituting from (1.6) for u_F in (1.7)

$$(a_x)_{\text{field}} = \frac{2u_D}{t} \tag{1.8}$$

and substituting in (1.8) for $(a_x)_{\text{field}}$ in (1.3),

$$\frac{2u_D}{t} = -\frac{e}{m} \cdot E_x$$

where u_D is the drift velocity in the x direction.

So
$$u_D = \frac{-eE_x t}{2m} \tag{1.9}$$

The ratio u_D/E is a constant and is called the *mobility* of the charge carriers, and is given the symbol μ (pronounced mu).

So
$$\mu = \frac{u_D}{E} = -\frac{et}{2m} \tag{1.10}$$

But, from eqn. (1.4), $J = -neu_D$.

So
$$J = \frac{ne^2t \cdot E}{2m} \tag{1.11}$$

where we have dropped the suffix x from E.

If this statement of Ohm's law is compared with the original

$$J = \sigma E \tag{1.1}$$

we see that

$$\sigma = \frac{ne^2t}{2m} \tag{1.12}$$

or, by using eqn. (1.10)

$$\sigma = -ne\mu \tag{1.13}$$

It is instructive to consider eqn. (1.12) in some detail. We shall rearrange it in the form

$$\sigma = (ne)\left(\frac{e}{2m}\right)t$$

The first term, ne, tells us that the conductivity is proportional to the amount of mobile charge per unit volume of material; the second term $e/2m$ shows, from eqn. (1.3), that the acceleration which an applied field can give to a charged carrier is directly proportional to the charge on the carrier, but inversely proportional to the carrier's mass. The last term, t, shows that the applied field can accelerate the carrier only during the mean free time, the next effective collision removing all trace of the drift velocity.

It is also clear from eqn. (1.12) that only two quantities distinguish one conducting material from another: n, the number of charged carriers per unit volume, and t, their mean free time. The former can be estimated fairly accurately, but determination of the latter is difficult. We shall show in the next chapter how Drude found a way around the problem.

QUESTIONS

1. Describe how a crystal of an ionic compound changes from insulator to conductor on melting.

2. Calculate the current flowing across 2×10^{-5} m^2 of iron (conductivity $1{\cdot}1\times10^5$ ohm^{-1} cm^{-1}) in an electric field of 0·3 V m^{-1} (66 A).

3. Explain what is meant by an isotropic material. Give two examples of isotropic, and two examples of anisotropic, materials.

4. "Heat is internal Kinetic Energy." Discuss and qualify this statement and explain briefly where the "heat" comes from when an electric current flows in a conductor.

5. What is meant by (a) the mean free time, and (b) the mobility, of an electron moving in a metal? Show how they are related.

CHAPTER 2

CONDUCTION AND HEAT

DURING the period between Gray's original work in 1730 and Drude's theory of 1900, a picture of electron conduction slowly emerged. A metal was assumed to consist of atoms each one of which contributes an electron to a common pool, and becomes a positive ion in the process. It was further assumed that the ions were held together by the attraction of the pool, and the attraction of the ions prevents the pool from escaping.

The electrons, when free, move about between the ions with random velocities, but when a potential difference is applied the electrons experience an overall motion one way or the other depending on the polarity, and this "drift velocity" is superimposed on the random motion (see Fig. 1.4).

What we have called a "common pool" of electrons is frequently referred to as the *electron gas;* of course, such a gas is very much condensed in comparison with (say) oxygen—indeed, the electrons in the "electron gas" may be 10 times closer than the molecules of oxygen at the same temperature and pressure. In this sense the electron "gas" may be thought of as more like a liquid. But here the resemblance ends; the charged electrons repel each other, whereas the neutral molecules of a liquid have quite a strong attraction for each other.

Initially, then, Drude pictured the negative charge of the elec-

trons and the positive charge of the ions as being "smeared" throughout the metal, neutralizing each other. He neglected the repulsions between electrons and the attractions between electrons and ions, and treated the electrons as uncharged particles (except for their function of transporting charge), and the ions as uncharged obstacles (except for their function of preventing the electron gas from escaping).

It was on this basis that Drude developed the ideas in Chapter 1; he soon realized the analogy between the electron gas and the kinetic theory of matter that had been developed during the previous two centuries by many investigations into the nature of heat and gases.

2.1. SOME KINETIC THEORY

About 1660 Boyle was investigating "the doctrine of the spring of the air", and his work was independently repeated a decade later by Mariotte who found that "air is condensed in proportion to the weight with which it is loaded". Gay-Lussac in 1825 extended this rule to all gases.

We can write Boyle's law as

$$P = \frac{\theta N}{V} \tag{2.1}$$

where P is the pressure of the gas, V its volume, N the number of molecules it contains, and θ is a temperature-dependent constant.

It is found that θ is directly proportional to the absolute temperature T of the gas concerned, and we say

$$\theta = kT \tag{2.2}$$

where k is a universal constant, called Boltzmann's constant, and equal to $1{\cdot}38 \times 10^{-23}$ joules K^{-1}.

As long ago as 1738 Bernouilli had adopted the now well-accepted physical model of a gas, in which the molecules are assumed to be in rapid random motion, and had deduced that the pressure of a gas is proportional to the square of the velocities of its molecules. It can be shown that

$$P = \frac{1}{3} m u^2 \frac{N}{V} \tag{2.3}$$

where m = mass of a molecule, and u = r.m.s. (root mean square) velocity of all the molecules.

Substituting for θ from eqn. (2.2) in eqn. (2.1) gives us

$$P = kT \cdot \frac{N}{V}$$

Comparing this with eqn. (2.3) we have

$$\tfrac{1}{3} m u^2 = kT \tag{2.4}$$

or

$$\tfrac{1}{2} m u^2 = \tfrac{3}{2} kT$$

But $\frac{1}{2} m u^2$ = kinetic energy of translation of a molecule, $= E_k$. So this simple picture *does* give the important result that the kinetic energy of the molecules is directly proportional to the absolute temperature. It is important to appreciate that this theory holds true only when the assumptions on which it is based are true, the most important of which is the assumption that the gas is "dilute" enough for us to neglect the actual volume of the molecules considered as solid lumps, compared to the volume which the gas as a whole takes up.

Gases which agree with this simple analysis are called "ideal" or "perfect" gases; deviations occur at high pressures and low temperatures.

2.2. THE MEAN FREE PATH

Drude extended the work of Chapter 1 by calling on the results of the Kinetic Theory. He tried to apply this theory to the electron gas in this way:

From eqn. (2.4) we can say that

$$u = \left(\frac{3kT}{m}\right)^{1/2} \tag{2.5}$$

where m, u now refer respectively to the mass and velocity of the electron in the electron gas, and T is the absolute temperature of the metal containing the electron gas. We cannot substitute this value for u directly into eqn. (1.2)

$$(\sigma = ne^2t/2m)$$

as u does not specifically appear in that equation for conductivity; but if we introduce the concept of "mean free path", represented by the symbol λ ("lambda"), we can then do so. We define λ by

$$\lambda = ut \tag{2.6}$$

Substituting eqns. (2.5) and (2.6) into (1.12) we have

$$\sigma = \frac{ne^2t}{2m} = \frac{ne^2}{m}\cdot\frac{\lambda}{2u}$$

$$= \frac{ne^2}{m}\cdot\frac{\lambda}{2}\left(\frac{m}{3kT}\right)^{1/2} \tag{2.7}$$

Drude's theory then predicts that the electrical conductivity of a metal should be inversely proportional to the square root of its absolute temperature.

2.3. THE JOULE HEATING EFFECT

It is of interest to note in passing that this simple theory gives the right form for Joule's law.

We saw from eqn. (1.9) that the average drift velocity is

$$u_D = \frac{-eEt}{2m}$$

So just before an effective collision the average final velocity is

$$u_F = \frac{-eEt}{m}$$

Between collisions, then, an electron gains a kinetic energy of

$$\frac{1}{2}mu_F^2 = \frac{1}{2}m\left(\frac{-eEt}{m}\right)^2$$

from the applied electric field. When the electron suffers an effective collision it must lose all this energy to the ion with which it collides. As the electron makes $1/t$ collisions per second, and there are n electrons per unit volume, the lattice of ions must gain a total energy per unit volume per second of

$$\frac{1}{2}m\left(\frac{eEt}{m}\right)^2 \cdot \frac{1}{t} \cdot n$$

$$= \frac{ne^2t}{2m} \cdot E^2 \text{ joules m}^{-3}\text{ sec}^{-1}.$$

But from eqn. (1.12) $\sigma = ne^2t/2m$. So the expression for the Joule heating effect becomes

$$\sigma E^2 \text{ watt. m}^{-3}$$

(The reader is probably familiar with this as V^2/R where this is just the electrical power dissipated by a p.d. of V applied across a resistance R. Note that the volume does not appear in this form. It can be shown that both forms are identical.)

2.4. THE FAILURE OF DRUDE'S THEORY

Unfortunately for Drude's theory, experiment does not bear out the relationship of eqn. (2.7); in fact, over a wide range of temperature, the electrical conductivity of a metal is found to vary inversely with the absolute temperature, *not* the square root thereof.

It seems reasonable to look at Drude's formula

$$\sigma = \frac{ne^2}{m} \cdot \frac{\lambda}{2} \cdot \left(\frac{m}{3kT}\right)^{1/2}$$

and see if any term there is temperature dependent. If n, the density of the free electrons in the metal, varied inversely with the square root of the temperature, we would have our required temperature dependence for σ. But, if n varies at all with temperature, it should increase with rising temperature, not decrease, as the increasing thermal agitation of the ions should shake free extra electrons.

We cannot find the required dependence with the mean free path either, because as a metal is heated it expands and the obstacles to the flow of electrons become farther apart, thus increasing rather than decreasing it. It is an interesting exercise to substitute typical values, say $n = 8{\cdot}5 \times 10^{-28}$ free electrons per m^3, and $\sigma = 64 \times 10^6$ mho m^{-1} at 273°K into eqn. (2.7), and estimate the mean free path λ. The reader would find that the mean free path for copper would be of the order of 42 nm. But

it is well known that the interatomic spacing is about 0·256 nm (from X-ray diffraction) (see Vol. 1, Sect. 3.1) and it is very hard to see why an electron should pass by about 42/0·256 = 164 obstacles before it makes an effective collision! Indeed this problem is even harder to understand when it is realized that increasing the pressure on a metal increases the conductivity, when Drude's theory indicates that this process would decrease the mean free path and therefore the conductivity.

The last difficulty with Drude's theory which we will mention here is the experimental fact that when a small amount of copper is added to another conductor metal (such as nickel), the electrical conductivity of the resultant alloy decreases, notwithstanding the fact that copper has a higher electrical conductivity than nickel.

Before we dismiss Drude's theory completely it is only fair to point out that it is very successful on one point: it had been noticed about 1850 that the ratio of the electrical to the thermal conductivity of any metal is approximately constant when they are measured at the same temperature, and this was first pointed out by Wiedmann and Franz. This seemed inexplicable until Drude suggested that both heat and electricity are carried by the electrons through metals, and used his theory to calculate the thermal conductivity of the "electron gas". He found that the ratio of the two calculated conductivities corresponded very closely with the ratio found experimentally.

2.5. THE WORK OF WIEN

In 1905 Lorentz reinvestigated the problem, using what is known as the "Boltzmann transport equation" and a simple model of the collisions between the free electrons and lattice

ions; but the use of classical statistics could not uncover the source of its failings. However, in 1913 Wien showed a brilliant insight; first, he suggested that the average speed of the electrons is independent of the temperature—that the electrons in the "electron gas" are not like the molecules in a normal gas in this respect; secondly, he brought the temperature dependence back into the problem by emphasizing that the ions in a metal are engaged in thermal vibrations, the amplitudes of which increase with temperature. Further, he suggested that the probability that an electron will make an effective collision with an ion is increased if the amplitude of vibration of the ion is greater, but the vibrational energy of the ions in a metal accounts for the heat content of the metal at that temperature, and it is well known that the heat content of a body is directly proportional to its temperature. So Wien was really saying that the probability of an effective collision between an electron and an ion is directly proportional to the temperature, and this gives the right temperature dependence to the electrical conductivity in Drude's formula.

It is interesting to note that Drude had put the responsibility for the temperature dependence of conductivity entirely on the electron gas. Wien ignored this and put the responsibility on the ions which Drude had ignored.

In 1928 Sommerfield recalculated the conductivities on the lines of Lorentz' theory, but replacing classical statistics by Fermi–Dirac statistics, thus putting Wien's suggestions on a strong theoretical basis. (The essence of Fermi–Dirac statistics is that not more than two electrons can occupy the same energy levels in one atom simultaneously. Classical statistics put no limit to the number in any level.)

2.6. THE MODERN PICTURE OF RESISTIVITY

We have seen that the electrons are supposed to move in the spaces between the ions, with a mean free path of the same order of magnitude as the separation of the ions. Figure 2.1 represents this classical model. It is well known, in fact, that the mean free path is several hundred times that given by this model.

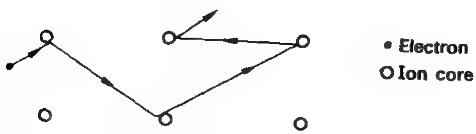


FIG. 2.1. The classical model of an electron moving through the lattice of ion cores

We have also seen that this simple model fails to account for the temperature and pressure dependences, and influence of impurities on the conductivity.

To develop an acceptable model we must use the wave motion aspect of an electron (Vol. 1, Sect. 2.9.3); because of this, electrons can pass through a *perfect* crystal without suffering any resistance at all, and this is due to the interference of the electron waves scattered by the periodic potential which composes the lattice of ion cores. There is an analogy here with the unattenuated passing of light through a perfect crystal. It follows too, that if all the ions were at rest, that the mean free path for electron scattering must be infinite. We must look to the deviations from the periodicity of the potentials in which the electrons move,

for the actual cause of the resistivity. Such deviations may be due to:

(a) vibrations of the lattice of ion cores;
(b) defects in the lattice structure of the crystal, such as vacancies, interstitials, and dislocations;
(c) the presence of foreign impurity atoms;
(d) boundaries, i.e. edges of the sample.

2.7. THE EFFECT OF TEMPERATURE ON THE CONDUCTIVITY

We are now in a position to understand more fully the effect (mentioned briefly in Vol. 1, Sect. 3.4) that temperature has on the electrical conductivity of conductors, semiconductors, and insulators.

We saw in eqn. (1.12) that

$$\sigma = \frac{ne^2t}{2m}$$

and we must now look for the explanation for the difference in conductivity between a good conductor and a good insulator, the factor being of the order of 10^{20}. As the charge e and the mass m of the charged carriers are the same, we must look for variations in n and t between the substances.

In a metal the number of electrons per unit volume is independent of T, but as T increases the amplitude of the atomic vibrations increases, so the scattering by lattice vibrations increases because of the more marked deviations from lattice periodicity. Consequently the mean free path decreases, and, as the velocity of the electrons is constant, the mean free time t and the mobil-

ity must decrease. So the conductivity decreases when the temperature of a metal is increased.

In an intrinsic semiconductor or insulator, the number of free electrons per unit volume n at any instant is dependent on the absolute temperature T and the ionization energy W_g (the amount of energy required to partly break a bond and release an electron). The relationship is

$$n = 5\times10^{21}\,T^{3/2}\exp\left(\frac{-Wg}{2kT}\right)\ \text{electrons m}^{-3} \tag{2.8}$$

The exponent in (2.8) provides the predominant part of the temperature dependence on n; the three-halves power term provides a relatively slow rate of change of n with T. It is clear from this and it is reasonable to expect, that n increases with temperature T and the less the ionization energy W_g. Consequently the conductivity increases when the temperature of an insulator or intrinsic semiconductor is increased. For silicon, $W_g = 1{\cdot}21$ eV and for carbon (in diamond form), $W_g = 5{\cdot}2$ eV; at room temperature of about 300°K the electrical conductivity of silicon is about 3×10^{-4} mho m^{-1}, whereas that of diamond is only about 10^{-14} mho m^{-1}. The use of this information for silicon will show, by substitution in (2.8) that n for silicon at room temperature is about 7×10^{16} electrons m^{-3}, whereas n for copper at the same temperature is about 10^{29} electrons m^{-3}, i.e. a factor of about 10^{12} greater. We do in fact find that the electrical conductivity of copper is about 59×10^{6} mho m^{-1}.

It is interesting to note in passing that Faraday, as early as 1833, noticed that the electrical conductivity of solid silver sulphide increased when he raised its temperature. He had discovered the first semiconductor!

EXPERIMENT 2.1: Conductance of materials, and the effect of temperature.

For this experiment a simple but sensitive meter capable of qualitatively comparing low conductances is needed. The circuit below (Fig. 2.2) is recommended.

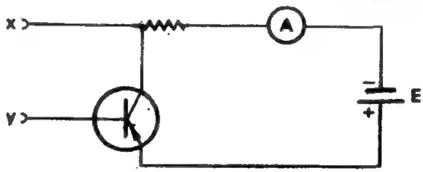


FIG. 2.2.

The transistor T is a type OC81; m is a 1 mA meter, and R a 1000Ω resistance. E is a 1·5-V dry cell. (The 'Unilab' 1 mA basic meter, with 1 V series resistance in position, is equivalent to R and m together.) The meter gives full-scale deflection when a resistance of about 10^5 Ω is connected across the clips, and correspondingly less for higher resistances. Care must be taken not to short the clips together; if this happens the transistor would suffer permanent damage after a few seconds. There will be a small reading on the meter even when the clips are open-circuited; this reading should be regarded as the "false zero".

(1) Hold one clip in each hand. Depending on the condition of the skin, for most people a full-scale deflection will be produced.

(2) Attach the clips to about 1 in. of writing paper; with them in position draw an increasingly thick pencil line to join the two clips.

(3) Connect a 220 k (220,000 Ω) carbon resistance between the clips. Heat it with a match, and note that the reading falls, showing that its resistance increases (conductance decreases) as temperature rises.

(4) Break the glass from a blown torch-bulb and detach the remains of the filament. This leaves two stiff wires joined by a glass bead. Connect the bulb to the clips. No deflection shows that the glass bead is an insulator. Now heat it with a match; suddenly its conductance will rise, but fall again as soon as the heat is removed. (How can you prove this is not due to conduction through the air ionized by the match flame, or through the soot deposited on the bead?)

(5) Connect a phototransistor type OCP71 to the clips, emitter to lead x of Figure 2.2, and collector to lead y. In average light full-scale deflection will result, but when the phototransistor is darkened by shielding with the hand, the deflection falls to about half-scale.

(6) Connect a specimen of pure germanium or silicon to the clips and observe the effect of heating with a match.

(7) Connect an ammeter (f.s.d. about 3 A) in series with an ordinary

rheostat of about 15 Ω and a 2-V accumulator. After checking that the design of the rheostat is such that it will not be damaged by being heated for a short while, play a bunsen flame on to the winding. Note the decrease in current, showing a decrease in conductivity as in part (3).

(8) Repeat (7), using a Radiospares thermistor type TH3 and an ammeter with a smaller f.s.d. between $\frac{1}{2}$ and 1 A. When this is heated the current will be seen to *rise*.

QUESTIONS

1. Describe briefly how electrons move in a metal (a) at room temperature, (b) at a dull red heat, and (c) at room temperature, when an e.m.f. is maintained across the metal.

Account for the increase in resistivity of copper when heated, and for thermionic emission.

2. Calculate the average kinetic energy of translation of molecules of gaseous oxygen at 100°C, given that Boltzmann's constant = $1{\cdot}38\times10^{-23}$ J K^{-1}.

What would be the result for hydrogen at the same temperature?

($7{\cdot}72\times10^{-21}$ J)

3. If Drude's original theory of conduction were correct, what would be the electrical conductivity of copper at 500°C, if its resistivity at 20°C is $1{\cdot}72\times10^{-6}$ ohm cm?

($3{\cdot}6\times10^{5}\ \Omega^{-1}\ \mathrm{cm}^{-1}$)

4. List and discuss experimental observations which do not bear out Drude's original theory of conduction.

5. Explain carefully why the resistivity of a metal rises when it is heated, whereas that of an intrinsic semiconductor falls.

6. The mobility of electrons was said (Chap. 1) to be determined by the frequency of effective collisions between electrons and atoms; why, in this "free-electron-gas" theory of conduction, is it unnecessary to consider the collisions of electrons with one another?

7. If electric conduction in a metal is considered, the free electrons have a resultant drift velocity in a certain direction. These electrons must exert a net time-average force on the atoms of the solid. If the material were free to move, then, would it do so, when an electric current passes?

8. At very low temperatures the specific heat of a metal is proportional to the cube of the absolute temperature. Assuming Wien's hypothesis is still valid, how should the electrical conductivity of a metal vary with temperature at similar low temperatures?

CHAPTER 3

SEMICONDUCTORS

3.1. HISTORICAL

The first recorded experimental work on semiconduction was done by Faraday in 1833 when he happened to pass an electric current through solid silver sulphide. He found the conductivity to be low, but to increase if he raised the temperature.

We showed in Section 1.2 how it is possible for an ionic crystal to conduct electricity by the small amount that it does, and from Section 1.3 on we explained how it is that metals conduct so well. The conduction of electricity in silver sulphide is odd, however, because this substance conducts much better than an ionic crystal but not nearly as well as a metal. This fact necessitates the introduction of a *third* group, semiconductors, in addition to metals and insulators.

As time went by several more members of this third group were found, and many strange properties were observed in them: for example, in 1873 it was found that when a light beam fell on selenium its electrical conductance increased. A year later several experimenters found substances which, when in contact with a metal, gave a resistance that did *not* obey Ohm's law. Although there was little physical understanding of the

mechanism of electrical conduction involved, some of the properties of these semiconductors were exploited technologically. The rectification of alternating current was achieved using copper oxide to copper rectifiers; and at lower power but higher frequency the "cat's whisker" detector—formed by having a sharp bit of metal wire in contact with a bit of carborundum, galena, silicon or tellurium—appeared in 1904 just as radio communication was beginning to develop. However, Fleming found the vacuum diode a far more reliable detector than the "cat's whisker" one, and the latter fell into disuse for some 30 years, and the thermionic valve had its hey-day.

But as telecommunications began to exploit higher and higher frequencies, the limitations of thermionic valves became apparent (see Vol. 1, Sect. 7.1) and during World War II, as radar was being developed, interest revived in silicon detectors and studies were made of the conduction mechanisms involved in semiconductors. A. Wilson, in 1931, had laid the foundations of a wave-mechanical theory that explained electronic conduction in pure and impure semiconductors, and this theory has been verified experimentally in more recent years. Not that experimental work on semiconductors is easy—for if a piece of material appears to be non-ohmic (i.e. not to obey Ohm's law), it may not be a bulk property of the material alone, but a surface property at the junction of the connecting wire with the material.

But by about 1935 it was clear that the latter possibility was correct, the non-ohmic behaviour arising at the junction.

After the war research really began seriously on semiconductors: as a starting point, as usual, the simplest elements were chosen. Silicon and germanium are both Group IV elements and crystallize, like the carbon atoms in diamond, by purely covalent bonding into simple atomic arrangements. In this chapter we go on to describe the present picture of semiconductors that has grown in the last two decades on the basis laid by Wilson; in the

next chapter we go on to describe the junctions that can be formed between semiconductor materials and which led to the transistor (or *trans*fer res*istor*) in 1948.

3.2. CLASSIFICATION OF MATERIALS

A useful chemical picture is given in Figure 3.1, in which the first ionization potential (i.e. the smallest energy required to remove one electron from the outermost occupied shell of an atom) has been plotted against atomic number. From the graph we see that the highest energies are required for the rare gases, and the lowest for the alkali metals. When this potential exceeds 10 eV the elements form insulating solids, and when less than 8 eV, conductors. But *some* conductors and *most* semiconductors fall between 8 and 10 eV.

Another viewpoint is to consider whether the conduction and valency energy bands are full of electrons, half-full, or empty, and the size of the energy gap between these bands. This subject was introduced in Vol. 1, Chap. 3, but we will repeat the important points here and develop it somewhat further. In Figure 3.2 the difference between a conductor, a semiconductor, and a dielectric (or insulator) is made clear by using energy band diagrams. Table 3.1 relates the electrical conductivity of various metals with the energy gap W_g.

TABLE 3.1

	Pb	Sn	Ge	Si	C (diamond)
Conductivity σ (mho m^{-1})	5×10^{6}	5×10^{5}	2·1	3×10^{-4}	10^{-14}
Energy gap W_g (eV)	0	0·08	0·75	1·21	5·2

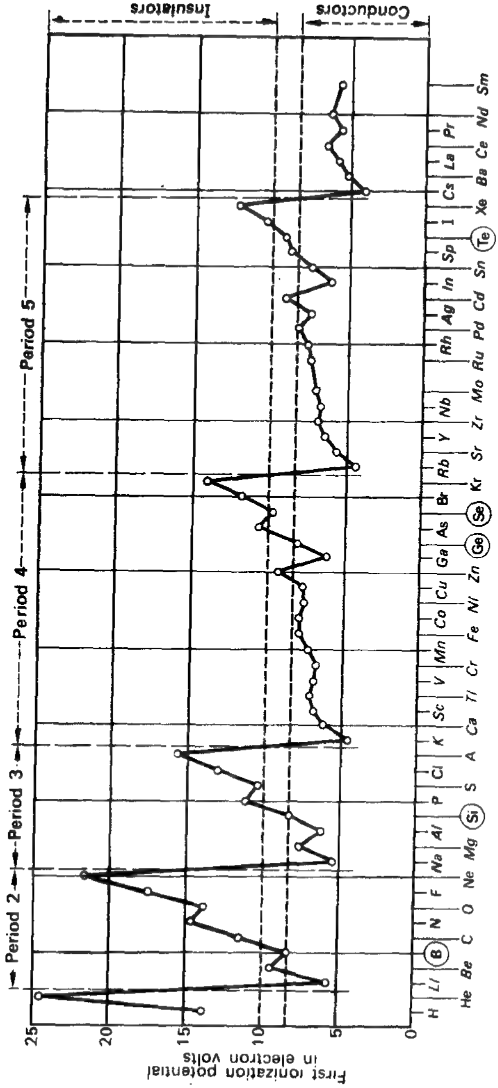


FIG. 3.1. Graph of first ionization potential against atomic number. (Courtesy of Alan Holden)

This table comprises elements in the fourth group of the periodic table, all of which crystallize in the diamond structure. It is clear that as W_g rises the electrical conductivity σ falls rapidly, because from eqn. (2.8), n (q.v.) falls rapidly, and $\sigma = ne^2t/2m$.

A semiconductor material is characterized by an energy band model in which the conduction band is empty, but the valence

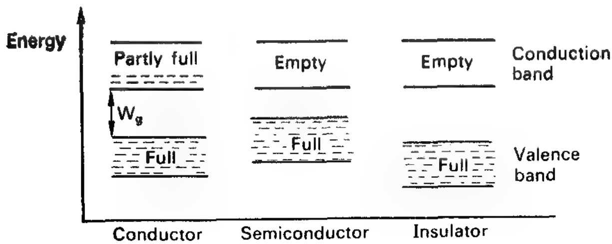


FIG. 3.2. The energy band diagram

band is full of the electrons, and the energy gap between these two bands is reasonably small, e.g. about 1 eV. (See Vol. 1, Fig. 3.5.)

In addition, one normally finds that the temperature coefficient of resistance is negative over at least part of its characteristic.

3.3. ELECTRONS AND HOLES

If a semiconductor is to conduct electrically at all its charged carriers must be able to gain energy from the applied field; in order to do so, electrons must leave the valency band, jump the forbidden energy gap, and arrive in the conduction band. If the temperature of the crystal is sufficiently high, and the

energy gap sufficiently small, thermal energy alone may be enough. Alternatively, an electron in the valency band might absorb enough light energy to hop across the gap. In either case once an electron gets into the conduction band it meets with a multitude of empty energy levels, and conduction can take place easily.

However, it is important to consider the empty levels left behind in the otherwise full valency band. Electrons in the valency band can move in response to an applied field by moving into these vacated levels, and can thus contribute to the resultant electric current. But the valency band is very nearly full of electrons and it is easier to keep track of this small number of empty levels than to study the motion of all the electrons in the band. These empty levels are called "holes", and because the motion of an empty level appears to be that of a positively charged electron, we can think of the holes as being positively charged—simply as a matter of convenience.

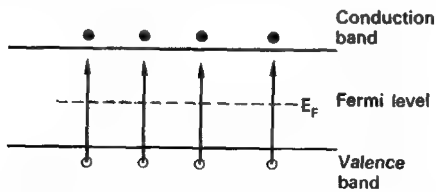


FIG. 3.3. Energy band diagram of an intrinsic semiconductor crystal

As every electron arriving in the conduction band must leave a hole behind in the valency band, the electron and hole densities are equal in a perfectly pure semiconductor crystal (see Fig. 3.3).

A free electron wandering about in the crystal may encounter a broken bond or hole, and recombine with it, and so two charge carriers would then be lost for the process of conduction. As

opposed to this recombination, absorption of energy by a bond from the lattice vibrations of the crystal could result in the generation of a hole–electron pair. At a given temperature the rates at which these processes occur are equal, and an equilibrium is reached, where the average number of electrons equals the average number of holes in the material. The process may be written as a "chemical reaction":

$$\text{one bound electron} \rightleftharpoons \text{one free electron} + 1 \text{ hole}$$

As the temperature is increased the average number of free charge carriers increases.

By considering the recombination process (say), an important principle can be realized: the probability that a particular free electron will encounter a hole in a given time interval must be directly proportional to the number of holes per unit volume that are available for recombination. But the total number of electrons that will encounter holes in this time interval is proportional to the probability for any *one* of them, multiplied by the number of free electrons per unit volume that are available for recombination. So the rate at which holes and electrons recombine is proportional to the density of holes multiplied by the density of free electrons. But, as we said earlier, at a given temperature the whole process is in equilibrium, recombination and generation occurring at the same rate. We must therefore conclude that

$$np = k \tag{3.1}$$

where n = density of free electrons, p = density of holes, and k is a temperature-dependent constant. Note that k does *not* vary with n and p, at a given temperature. Equation (3.1) is known as the Law of Mass Action.

The conductivity of the semiconductor is dependent on the

hole–electron densities and, as we saw in eqn. (2.8), these are governed primarily by the energy gap. But this gap is an inherent property of a particular crystal, and that is why the word *intrinsic* is used to describe such semiconductor crystals. If there is any deviation from perfect lattice periodicity in the crystal, e.g. due to foreign "impurity" atoms, the intrinsic properties of the crystal will be considerably modified. Indeed, it is the *deliberate* addition of chemical impurities to intrinsic semiconductors, with the consequent change of electrical properties, that is so very important—as we shall see in the next section.

3.4. EXTRINSIC OR IMPURITY SEMICONDUCTORS

We showed in Vol. 1, Sect. 1.9 the effect on the crystal structure of a group IV element (e.g. silicon or germanium) of a group 3 or group 5 impurity (boron or phosphorus, for example). (Those readers without access to Volume 1 will find a brief summary in Appendix VII.) The intrinsic semiconductor would originally have a purity of about 1 in 10^{10}, and to this is added about 1 part in 10^7 of a selected impurity. We shall see later how these extreme purities are obtained. The reader will appreciate that the addition of boron to silicon (say) results in a number of holes in the crystal structure; consequently this impurity is called an "acceptor" (because it requires, or accepts, an electron in recombination) and the semiconductor is called "*p*-type". Conversely, the phosphorus impurity has one electron too many and is called a "donor" (because it donates its extra electron to the conduction process) and the semiconductor is called "*n*-type". In this section we will be concerned with the effect on the energy band model of the deliberate addition of impurities. This effect is shown in Figure 3.4. It requires very

little energy for an electron to leave the valence band of the silicon and move up into the "acceptor energy level", which represents the holes contributed by the (say) boron impurity. This leaves a hole in the valency band. The holes so formed are *not* associated with any electrons in the conduction band. Conversely, if the added impurity were phosphorus (say), thermal

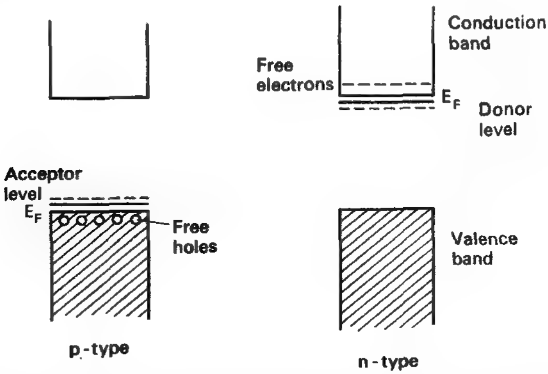


FIG. 3.4. Energy band diagram of an extrinsic (impure) semiconductor crystal

energy might be quite sufficient to raise the free electron from the "donor energy level" into the conduction band of the silicon. There is not an equal number of associated holes in the valency band, and conduction this time takes place because of the electrons in the conduction band. This, of course, is why it is referred to as "*n*-type". Because the valence band levels in "*p*-type" crystals have more holes, and similarly because the conduction band levels in *n*-type crystals have more electrons, than they would if intrinsic, these types of semiconductor are called ***extrinsic.***

If both *p*-type and *n*-type impurities are present, the resultant number of charge carriers is simply the difference between the

numbers of acceptor and donor atoms available: if these concentrations were the same the material would behave as if intrinsic, and this fact can be utilized to produce intrinsic material of a purity greater than that available by purefaction technology. The extremely small concentration of acceptor and donor impurities can fortunately be found quite easily and their

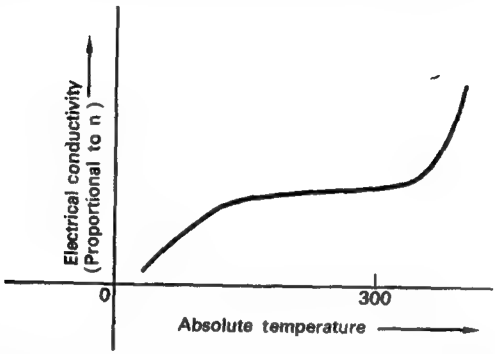


FIG. 3.5. Graph of the electrical conductivity of an intrinsic semiconductor against the absolute temperature

identity established by the Hall effect, which is discussed in Section 4.4.

Figure 3.5 shows the variation of conductivity of an extrinsic semiconductor with temperature. Near absolute zero the impurities are inactive because the donor electrons are bound to their respective donor atoms, so very few hole–electron pairs are created. But as the thermal energy increases the impurities become active and, on the linear portion, are eventually fully activated, each donor atom producing one electron. These electrons cannot jump the energy gap and the material behaves extrinsically. Eventually at some temperature greater than room temperature, thermal generation begins, electrons jumping the gap into the conduction band. As this process increases there will exceed the

number given by the donor atoms, and the material behaves intrinsically. It is this process which often mars the performance of extrinsic semiconductors at high temperatures.

Of course impurities other than boron and phosphorus can be added: aluminium, gallium and indium are common acceptor

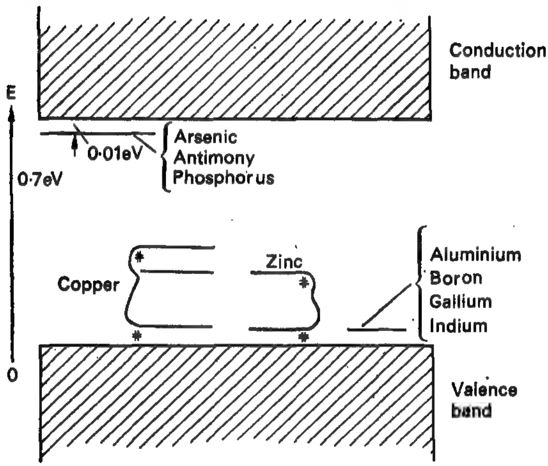


FIG. 3.6. Creation of energy levels in the "forbidden gap" of a semiconductor crystal

impurities, whilst arsenic and antimony are other donors. If, however, copper or zinc are added, triple and double impurity levels respectively would be established. This is because copper, for instance, having only one electron in its outer shell, can accept another three, thus giving rise to three acceptor energy levels. Figure 3.6 represents this for germanium. It is clear from Figure 3.6 that the commonly used impurities have energy levels lying very close to their respective bands—typically 0·01 eV. Obviously it would take considerable energy to excite a carrier from one of the copper energy levels.

This brings us to one final important point: impurities that are present in the crystal, which establish energy levels in the middle of the energy gap, *must* be removed. We need the life-time of the electrons and holes to be as large as possible, and theory shows that in perfectly pure germanium it is about 1 sec, i.e. on the average an electron in the conduction band exists for a second before recombining with a hole in the valence band. But if about 1 part in 10^8 of copper is present, the life-time falls to about 10 μsec, the three acceptor levels of copper acting as a kind of "step ladder" letting the electrons down into the valence band.

Not only must our semiconductor material be intrinsically very pure indeed, but the atoms in the crystal must be properly arranged—if this is not so, energy levels are again created in the middle of the energy gap. We next describe how extremely pure single crystals are produced, and discuss their uses.

QUESTIONS

1. If an electron moving round the nucleus in an atom has kinetic energy, and the same electron from the atom and at rest has none, why is energy needed to remove an electron?

2. The total number of charge carriers in a piece of pure germanium is 20×10^9 at a certain temperature. If a piece of germanium is made of the same size and it contains 21×10^9 arsenic by way of impurity, how many charge carriers will this specimen have at the same temperature?

3. Under what circumstances can the addition of impurities to a semi-conductor *decrease* its electrical conductivity.

4. Discuss briefly the significance of the term effective mass in the context of electrical conduction in crystals. How does the motion of an electron in a solid differ fundamentally from that of an electron *in vacuo* and what is the meaning of the term mobility?

Derive an approximate expression for the mobility of an electron in the conduction band of a semiconductor in terms of the mean free time

between collisions with lattice atoms. Why is such a simple derivation unsatisfactory in the case of a degenerate electron gas?

Describe experiments to measure (a) the majority carrier density and (b) the majority carrier mobility in a sample of semiconductor material.

(I.E.E. Part 3, Physical Electronics, June 1967)

5. Crystalline pure germanium has $4{\cdot}5 \times 10^{28}$ atoms m^{-3}. At 300°K one atom in 2×10^{9} is ionized. The mobilities of electrons and holes at 300°K are respectively $0{\cdot}4$ m^2/V s and $0{\cdot}2$ m^2/V s. Determine the conductivity of pure germanium.

Estimate the conductivity of germanium by the addition of 1 part in 10^7 of a trivalent element, at 300°K.

(U.L. B.Sc. (Eng.) III, Electrical, 1966)

6. Estimate the minimum percentage of donor impurities in a silicon sample that are ionized at room temperature, if the donor levels are about $0{\cdot}05$ eV below the bottom of the conduction band. Assume that the density of donors is less than 10^{-3} of the effective density of states in the conduction band, and that the density of holes in the valence band is small.

(C. of E.I., Electronic Engineering, Dec. 1966)

7. Explain the fundamental difference (a) between a conductor and an insulator, (b) between an extrinsic semiconductor and an intrinsic semiconductor. Illustrate your answer with energy-band diagrams. How does the conductivity vary in each case as the temperature is increased?

Calculate the minimum conductivity of silicon if its electron mobility is $0{\cdot}12$ m^2/V s and its hole mobility is $0{\cdot}05$ m^2/V s. Intrinsic silicon at room temperature has a carrier density of $1{\cdot}4 \times 10^{16}$/m^3.

(I.E.E., Part 3, Nov. 1967)

CHAPTER 4

SEMICONDUCTOR TECHNOLOGY AND SIMPLE DEVICES

4.1. TERMINOLOGY

There are several words peculiar to semiconductors that need explanation. The process of adding impurities to a semiconductor crystal is called *doping*, and the impurities so added may, like boron, contribute holes (i.e. "accept" electrons), in which case the boron atoms are called *acceptor* impurities; or, like phosphorus, they may contribute electrons, in which case the impurity atoms are called *donor* impurities. In an extrinsic semiconductor the hole and the electron densities are not necessarily equal; the more plentiful charge carriers are called the *majority* carriers, and the others *minority* carriers. So a *p*-type semiconductor has holes at its majority carriers, whilst an *n*-type material has electrons as the majority carriers. We shall see in the next chapter that a *junction* is formed by placing these two types of semiconductor very close together.

4.2. CRYSTAL PREPARATION

We saw in the last chapter that the semiconductor must not only be intrinsically pure to about 1 part in 10^{10}, but it must also be effectively one large perfect crystal. It is indeed remarkable that these conditions can be met. Chemical refining can reduce the impurity content to about 1 part in 10^{6}, and we must use

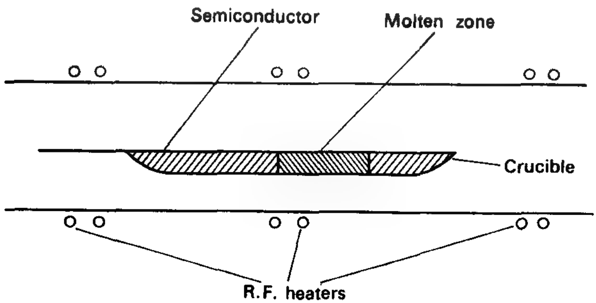


FIG. 4.1. The zone refining process

a metallurgical technique to reduce it further. A popular way is by *zone refining*, illustrated in Figure 4.1.

In this method the long impure bar of semiconductor is loaded into a crucible and this is moved through a series of R.F. heating coils, of sufficient power to melt the semiconductor (950°C for germanium, 1410°C for silicon). The impurities present have a tendency to remain in the molten zone so, as the molten zone moves through the bar, the impurities collect at one end. The bar can be passed through the system several times; eventually the bar is removed and the impure end cut off and discarded.

Having attained an extremely high level of purity in this way,

we now need it in perfectly crystalline form; for this we use a *crystal puller* (see Fig. 4.2).

The purified semiconductor is placed in the crucible, and when molten is referred to as the "melt". The melt is kept just above the melting point. A small seed crystal of the same material

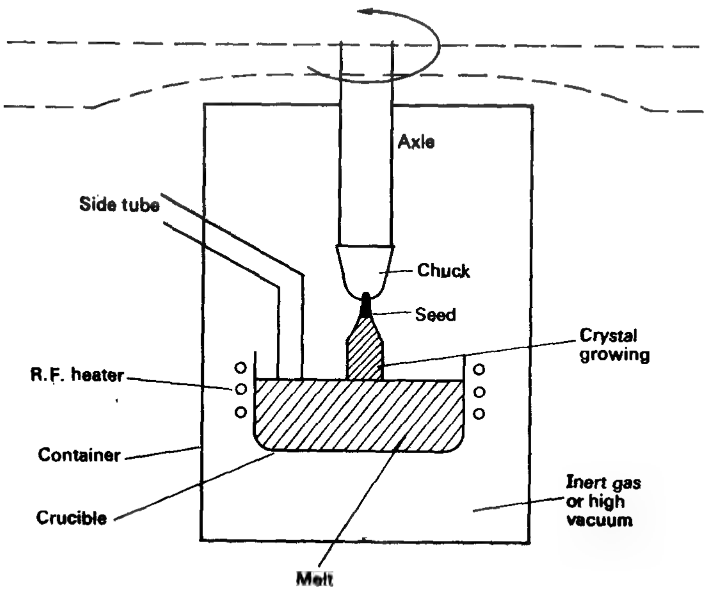


FIG. 4.2. A crystal puller

as the melt is then introduced into it, and withdrawn at about 1 mm/min; at the same time the chuck holding the seed rotates at about 5 rev/min, to stir the melt. As the seed is withdrawn the liquid melt freezes onto it, the deposited atoms following the crystal structure of the seed. Thus a pure and perfect crystal is grown. The heating is usually by a radio-frequency heating system running at perhaps 10 MHz and carrying a current of 100 A; some protection for the operator is therefore usually

necessary. To reduce contamination of the melt the apparatus can be enclosed in an inert atmosphere or high vacuum.

Crystals of *p*- or *n*-type semiconductor can be produced simply by using the respective semiconductor as the melt and seed. A crystal which changes from one type to the other along its length can be made by dropping the relevant dope in the form of a pellet down the side tube at the right time.

The zone refining process can also be adapted to grow crystals and add impurities.

4.3. THE MATERIALS IN USE

The group IV elements silicon (Si) and germanium (Ge) are by far the most common in general use, because practical devices, of which the transistor is only one, are fairly easily constructed from them, and because their intrinsic properties are good. From Table 4.1, because Si has the greater energy gap, its conductivity is less than that of Ge at the same temperature. The bonding of the electrons in an Si atom is much stronger than in a Ge one, and silicon therefore has fewer electrons and holes intrinsically available at room temperature. Of the two, Ge is the easier to prepare because a greater degree of impurity is acceptable; an Si crystal would go extrinsic much sooner than a Ge one, for the same impurity density.

Carbon, too, in the diamond form, is a Group IV semiconductor, but rather expensive to work with, as well as having a very high melting point.

There is no reason why a semiconductor must be an element, as is the case with the three foregoing examples. Compound semiconductors, such as metallic oxides and sulphides, are becoming of increasing importance in practical devices.

TABLE 4.1

Material			Energy gap (eV)	
Elements	Carbon (diamond)	C	5·3	
	Germanium	Ge	0·72	
	Silicon	Si	1·1	
Compounds	Cadmium telluride	CdTe	1·45	Photoconductors
	Cadmium selenide	CdSe	1·74	
	Cadmium sulphide	CdS	2·45	
	Gallium arsenide	GaAs*	1·34	
	Gallium phosphide	GaP*	2·25	
	Indium antimonide	InSb*	0·18	Infra-red detectors
	Lead selenide	PbSe	0·27	
	Lead telluride	PbTe	0·33	
	Lead sulphide	PbS	0·37	
	Magnesium oxide	MgO	7·3	
	Zinc oxide	ZnO	3·3	

* Intermetallic compounds, viz. compounds whose constituent elements are metals.

4.4. PHOTOCONDUCTORS

If a photon of light incident on a material has energy greater than the forbidden energy gap, it may raise an electron from the valency band into the conduction band (Fig. 4.3). Consequently the conductivity of the material increases (see Expt. 4.1). In practice one uses a material whose conductivity is as near zero as possible when in the dark, so that the reception of a photon causes the greatest percentage change in conductivity (see Vol. 1, Sect. 11.2.3). The compound semiconductors CdTe, CdSe, and CdS are very suitable. Taking CdTe as an example,

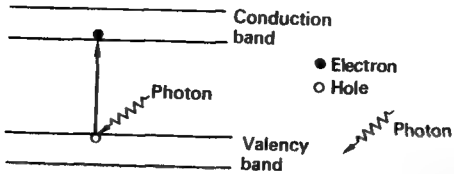


FIG. 4.3. Creation of a hole-electron pair by photon bombardment

we see from Table 4.1 that the energy gap is 1·45 eV $= 1{\cdot}45 \times 1{\cdot}6 \times 10^{-19}$ joule. If, then, a photon arrives with energy greater than this, conduction occurs. The energy of a photon is given by $E = h\nu$ where h = Planck's constant, $6{\cdot}625 \times 10^{-34}$ joule-sec, and ν, the frequency of the light, is equal to c, its velocity divided by wavelength λ.
So

$$1{\cdot}45 \times 1{\cdot}6 \times 10^{-19} = 6{\cdot}625 \times 10^{-34} \times \frac{3 \times 10^{8}}{\lambda}$$

because c, the speed of light, is 3×10^{8} ms^{-1}.
Therefore

$$\lambda = \frac{6{\cdot}625 \times 10^{-34} \times 3 \times 10^{8}}{1{\cdot}45 \times 1{\cdot}6 \times 10^{-19}}$$

$$= 860 \text{ nm } (1 \text{ nm} = 1 \text{ nanometre} = 10^{-9} \text{ metre}).$$

Cadmium telluride would therefore be suitable for use with light of wavelength 860 nm (or less), which is in the infra-red region. Similarly it can be shown that CdSe and CdS give maximum response to 710 and 530 nm respectively, the latter being in the green part of the visible spectrum.

The long-wavelength cut-off of the material's response is, of course, due to the inability of less energetic (longer wavelength) photons to push electrons across the energy gap. A short-wavelength cut-off occurs because short-wavelength light is strongly absorbed in passing through the crystal, thus activating the surface alone.

As an increase of electrical conductivity occurs when the crystal is illuminated, this can be used to trigger an electronic circuit. Doors can be opened automatically when a person interrupts a beam of light; street lights can be switched on automatically at sunset; objects passing on a conveyor belt can be counted; photographic exposure times can be accurately determined. Aeroplanes can be detected hundreds of miles away, by the intense radiation in the far infra-red coming from their exhaust systems. A material with a small energy gap, cooled in a liquefied gas to reduce the intrinsic conductivity, is used for this. Electrostatic photocopiers use zinc oxide which retains an electrostatic charge quite well in the dark, but whose conductivity rises, allowing the charge to escape, when illuminated; a permanent print is made from the "charge picture".

4.5. THE HALL EFFECT: MAGNETOMETERS

In 1879 E. H. Hall discovered the effect which now bears his name. He was working with gold foil and found that when it was carrying a current along it, and a magnetic field was applied perpendicular to the direction of the foil, a very small voltage was developed perpendicularly across the foil. This is represented in Figure 4.4.

A current I flows into the specimen from the left; a magnetic field of flux density B is applied perpendicularly to the major face of the specimen (whose length is l, width w, and thickness t). V_H is the Hall voltage that is established in the third dimension. We shall derive a simple equation for this voltage.

Consider an electron moving from right to left in Figure 4.4 (i.e. in the opposite direction to the conventional current flow I). It experiences a force in the direction shown, of magni-

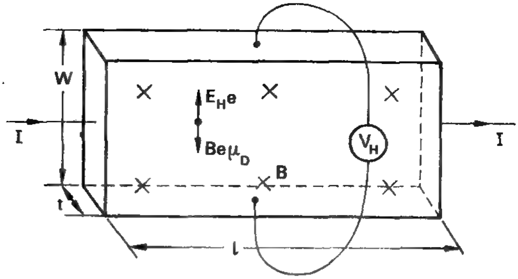


FIG. 4.4. The Hall effect

tude Beu_D where e is its charge and u_D its drift velocity. Consequently it moves in the direction of this force. Many electrons, of course, do this, and one side of the specimen gets a surplus of electrons whilst the other side gets a deficiency. So a charge gradient exists across the specimen, and an electric field E_H is established. But this exerts a force E_He on each electron, and an equilibrium is established for which

$$Be\mu_D = E_He \tag{4.1}$$

From eqn. (1.4) current density $J = neu_D$

and
$$J = \frac{I}{wt} \tag{4.2}$$

But
$$E_H = \frac{V_H}{w} \tag{4.3}$$

from the definition of electric field intensity as the potential gradient.

From (4.1) and (4.2) we get

$$\mu_D = \frac{J}{ne} = \frac{I}{newt} \tag{4.4}$$

Substituting for E_H, u_D from (4.3) and (4.4) into (4.1) gives

$$Be\frac{I}{newt} = \frac{V_H e}{w}$$

and

$$V_H = \frac{1}{ne} \cdot \frac{BI}{t} \tag{4.5}$$

From (1.13), $\sigma = -ne\mu$

so eqn. (4.5) becomes

$$V_H = -\frac{\mu}{\sigma} \cdot \frac{BI}{t} \tag{4.6}$$

The ratio of the mobility of the charge carriers μ to the electrical conductivity σ has a special significance and is called the Hall coefficient, R_H. So we have

$$V_H = R_H \cdot \frac{BI}{t} \tag{4.7}$$

From (4.7) we see that if the thickness t of the specimen is known, and the value of B (the flux density) and I (the current), measurement of the Hall voltage V_H enables the Hall coefficient to be found. As this is the reciprocal of the charge carrier density ne, we now know the density of the extremely small amounts of impurities that have been used to dope the intrinsic semiconductor. But there is even more to it: with a p-type semiconductor the Hall voltage has the *opposite polarity* to that obtained by an n-type material, and the interpretation must be that, whereas in n-type material conduction is by electrons, in p-type material conduction takes place by positive holes. This is very strong evidence for the reality of "hole conduction".

Of course, if in eqn. (4.7) the values of t, I, and R_H are known, measurement of V_H enables the magnetic flux density B to be found. The device thus makes a very convenient magnetometer, with the advantage that the specimen does not have to be jerked

out of the field, as in the traditional search-coil/ballistic galvanometer magnetic flux meter. Instead, a steady Hall voltage is developed, despite the absence of movement of the "probe" relative to the field. Experiment 4.5 shows the use of a Hall probe as a fluxmeter; the value of the Hall voltage is typically several millivolts for a specimen used with reasonable values of thickness, current, and field. Indium arsenide is favoured as the specimen, because of its high carrier mobility.

4.6. THERMOELECTRIC SEMICONDUCTORS

It is found that if a piece of *p*-type germanium is connected to a galvanometer by two metal connections, and one side of the germanium is heated, the galvanometer indicates that a current flows. If *n*-type germanium is used, the galvanometer deflects in the opposite direction. This thermoelectric effect was discovered first by Seebeck during the last century, whilst experimenting with junctions between dissimilar metals, and is named after him.

The effect is reversible—an electric current can be caused to flow through the metal contacts, and a temperature difference will be established across the semiconductor. This is called the Peltier effect, and is revolutionizing refrigeration because cooling is thereby produced without motors and compressors (see Expt. 4.2). We shall now show how it is possible for an electric current in a semiconductor to "pump" heat from one place to another.

In Figure 4.5 we show a piece of *n*-type semiconductor sandwiched between two metal contacts, across which a p.d. is applied, to form what is called a "frigistor".

Electrons flow from the negative pole of the battery into metal

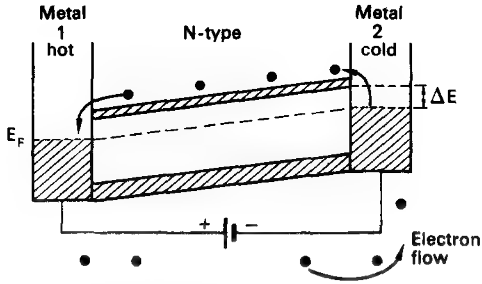


FIG. 4.5. The semiconductor thermoelectric effect

2, and from metal 1 back to the positive pole. We have shown the energy bands of the semiconductor tilted to account for the connection of the battery, for there is little energy decrease across the metals (because their conductivity is so high), and we show their energy bands not tilted.

Consider electrons in metal 2. The most energetic of them (i.e. the "hottest") can jump the energy gap into the semiconductor, "roll down" its conduction band energy level, and pass eventually into metal 1. So the most energetic electrons are taken from metal 2 to metal 1, thus metal 2 is continuously losing energy and gets progressively colder. The cooling effect is proportional to the current flowing, but, of course, there is always the Joule heating effect, which is proportional to the *square* of the current. Consequently, the temperature difference created reaches a maximum value for a certain current, and decreases with higher currents. Bismuth telluride, for example, produces a maximum temperature difference of about 45°C at a current of about 10 A.

A combination of *p*-type and *n*-type semiconductors can be used in series. As holes and electrons have opposite charges, heat is "pumped" in opposite directions in the two types, so the cooling effects (from the *p*/*n* junction) are additive. Temperature differences up to 80°C can be obtained in this way.

There are also practical devices based on the Seebeck effect, in which a temperature difference across a semiconductor is used to produce a flow of electric current.

4.7. THERMISTORS

These devices are sometimes called NTC resistors, because they have a *n*egative *t*emperature *c*oefficient of resistance. They are usually made of metal oxides, and have a range of values from about 1 Ω to several megohms; but whatever the nominal value, it *decreases* as the device is heated. The rate of change of resistance is large, making the thermistor capable of measuring changes in temperature of the order of 10^{-6} °C. This is reasonable, because in a semiconductor the number of electrons excited across the forbidden energy gap is strongly temperature-dependent. Not only is the thermistor used widely in industry for temperature measurement and control, but it finds many applications in electronic circuits in general, where its negative temperature coefficient can be used to counteract the positive temperature coefficients of ordinary resistors, valve heaters, cathode-ray tube deflection coils, etc.—so that, for instance, a television picture remains stable as the set gets warmer (see Expt. 4.3).

4.8. STRAIN GAUGES

We mentioned in Section 2.4 that one of the failures of Drude's theory was the fact that decreasing the pressure on (i.e. stretching) a piece of metal increases the resistance, whereas his

theory predicted the opposite. This effect is quite pronounced in silicon; a 1% change in resistance is caused by a 0·1% change in length. A simple explanation is that when silicon is stretched, the Si atoms get further apart and the forbidden energy gap increases. So the hole–electron density falls, and so does the conductivity. The method of use briefly is this: strain gauges are fastened to the specimen being investigated, e.g. a girder in a bridge structure, and various loads applied. The strains produced (change in dimension compared to original value of dimension, usually length) may be calculated from the change in resistance of the gauge, or may be read directly from a previously calibrated meter. These are compared with the maximum permissible strains for the material under investigation. The effects of temperature changes must be eliminated, and this may be done by "balancing out" in an electrical bridge circuit (see Expt. 4.4).

EXPERIMENT 4.1: Photoconductivity.

This is easily demonstrated using the Unilab cadmium sulphide (CdS) cell, which is supplied mounted in series with a 1·4-V pen-torch battery. Connect the terminals to a 50-mA meter. Note that the dark current is 0, and in strong daylight full-scale deflection may be attained.

Calculate the effective resistance of the CdS cell in bright daylight.

Further possibilities. In a darkroom, find how the resistance of the CdS cell depends on intensity of illumination, assuming that the light from the filament of a 12-V 6-W lamp obeys the inverse square law.

EXPERIMENT 4.2: To demonstrate the semiconductor cooling (Peltier) effect.

Apparatus required: Nicolson Peltier Demonstration apparatus, or Proops "Thermo-Pet" demonstration.

Procedure. Using the Nicolson apparatus, connect the terminals marked "galvanometer" to a lamp-and-scale galvanometer, and set the instrument to its most sensitive range. Touch the thermocouple for an instant with a finger, and note that the slight heating sends the spot off the scale. Now connect the bismuth telluride junction to a single 2-V lead–acid accumulator (or "nife" cell) with the polarity as marked. About 12 A are taken, so short

and fairly thick leads are needed. For a short time the cooling effect dominates, and the spot on the galvanometer is deflected in the opposite direction, but the Joule heating effect soon takes over, and the spot then moves in the same direction as when the thermocouple was touched with a finger.

Reverse the connections to the Bi–Te junction and note that the galvanometer indicates heating only. Do not leave on for more than a few seconds.

The Proops demonstration comprises several junctions in series, and will work from two fresh U2 cells in parallel. Wet the top surface of the "frigistor". When current is passed one way, a film of ice forms; when the current passes the other way, after a moment steam is seen to rise.

EXPERIMENT 4.3: Temperature resistance characteristic of a thermistor.

Apparatus: 12 V D.C. supply, up to 5 A.
Rheostat, 15 Ω 5 A.
Voltmeter, f.s.d. 3 or 5 volts (V).
Ammeter, f.s.d. 5 A (A).
Thermistor type TH5 ("Radiospares").
Boiling tube, thermometer (0–100°C), liquid paraffin.

Procedure. Connect the thermistor in the circuit below (Fig. 4.6) and twist its leads to support it against the thermometer bulb. Lower thermistor and thermometer into boiling tube, and cover with about $1\frac{1}{2}$ in. of liquid paraffin.

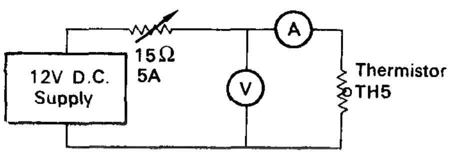


FIG. 4.6.

Adjust rheostat until voltmeter reads 1 V; quickly read temperature and current. (Readings will be changing as liquid is heated by thermistor dissipation.) Advance current to 3 to 4 A, and allow temperature to rise to about 30°C; reduce until voltmeter again reads 1 V, and quickly take new temperature and current readings. Continue, letting thermistor itself heat the paraffin, to 100°C if possible. The resistance will be found to vary from about 4 Ω at room temperature to about 0·2 Ω at 100°C. Plot a graph of resistance against temperature.

The rated dissipation of the thermistor is 1 W, but as it is surrounded by an effective "heat sink" (the liquid) in this experiment, many times this dissipation will not cause damage. Bubbling around the thermistor indicates

that heating is too rapid; currents in excess of 3 A for any length of time may damage the thermistor by intense local heating.

Problems. (a) Comment on the idea of using thermistor in a suitably calibrated circuit as a temperature measuring device.

(b) Suggest situations in which the thermistor, with its negative temperature coefficient, might be used to protect other items in an electronic circuit against overload.

EXPERIMENT 4.4: To demonstrate the operation of a strain gauge.

Messrs. Proops Brothers Ltd. can supply a suitable element. It is *very* fragile, and can easily be broken in being taken out of the cellophane packet. So the first thing to do is to secure it to the mid-point of a half-metre rule.

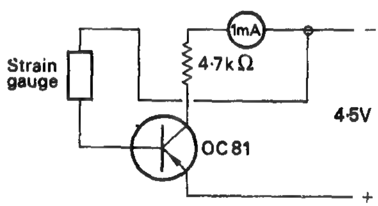


FIG. 4.7.

Lay it flat in the middle, and carefully but firmly stick it down with half-inch self-adhesive masking tape (Sellotape could be used).

Fix two flexible leads to the rule, then solder their ends to the foil contacts of the gauge, in such a way that they exert as little as possible force in any direction on the element.

Connect the gauge in the circuit of Figure 4.7. When the rule is bent along its length from the ends, the meter reading will vary by an amount which can be read quite easily, before returning to its former value. When the rule is released, the reading temporarily changes in the opposite direction. If the rule is bent in the opposite direction, this also reverses the directions of the reading changes. Note that twisting the rule produces only a marginal effect.

Clamp one end of the rule to the bench top, so that it is horizontal. At or near the other end fix a sling or scale-pan so that loads of up to 20 kg can be easily and rapidly added or removed. Tabulate the changes of reading produced by adding or removing loads of up to 20 kg, and graph load (which is proportional to strain for small values) against change of reading produced.

Further projects. Devise a simple circuit which reads 0 until a strain change occurs. Then devise a circuit which will produce a larger deflection for the same strain change. (This might be left until Chapter 10 has been absorbed, or the student is more familiar with designing transistor D.C. amplifier circuits.)

EXPERIMENT 4.5: The Hall effect, and its use in a fluxmeter.

Set up the circuit of Figure 4.8, in which the 25 kΩ potentiometers VR_1 and VR_2 are used to zero the galvanometers when the germanium chips are in zero (or earth only) field, by balancing out random voltage gradients.

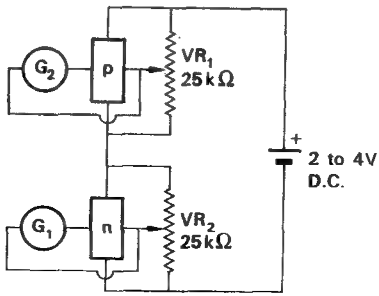


FIG. 4.8.

Ensure that the two galvanometers, if their terminals are not distinguished by + and − signs or colour, will each deflect in the same direction when a current is passed in the same way. Some centre-zero galvanometers have randomly connected terminals.

Place each end in turn of an "Eclipse" bar magnet over each chip. Observe the opposite deflections produced, and use the left-hand rule (see text) to determine or check the type of carrier (hole or electron) which is dominant.

("Unilab" supply a perspex "biscuit" containing a germanium slice (*p*- or *n*-type) and potentiometer, ideal for both parts of this experiment.)

To see how the effect is used in a fluxmeter, it is necessary to have a flux source which can easily be quantitatively varied. Anything wound on a ferromagnetic material may give misleading results, because of the hysteresis effect, so the 300-turn coil of a demountable transformer, without core, is ideal. It will pass about 9 A at 18 V, but should be restricted to about 30 sec of working at currents in excess of the rated 4 A. The most suitable supply is the 0–25-V A.C./D.C. "Variac" kind. With this arrangement, flux may be assumed proportional to current.

Connect the coil, an ammeter (0–10 A), and the supply in series, and place the chip centrally over one end. With the coil supply off, zero the

galvanometer using the potentiometer of either of the demonstration circuits. Take readings of galvanometer deflection against coil current. A straight-line graph will indicate that the deflection of this fluxmeter is proportional to the flux.

QUESTIONS

1. Describe carefully the distinction between "acceptor" and "donor" impurities, and give two examples of each. State from which of the groups of chemical elements they come, and why.

2. Describe the process known as "zone refining", and give examples of the degree of purity which can be attained by this method.

Why are silicon and germanium used more than any other element in semiconductor manufacture?

3. What is meant by the energy gap of a semiconductor?

The energy gap of gallium phosphide is 2·25 eV. What is the longest wavelength of incident electromagnetic energy which could promote conduction?

Why has lead selenide been used for so long as an effective photoelectric material?

4. A Hall effect experiment is performed on a rectangular block of copper 0·1 m long (in the direction of the current), 0·001 m thick (in the direction of the flux density B), and 0·01 m wide. Hall potential connections are made to the narrow sides. For a current of 40 A and for a flux density $B = 1{\cdot}5$ T, the difference in Hall p.d. on reversal of B is 6·6 μV. Indicating the relevant theory, calculate for copper: (a) the Hall constant, (b) the carrier density. The mobility is 0·0032 m^2/V s.

(C. of E.I., Electronics, Dec. 1966 amended)

5. Describe an experiment for determining the sign and density of the majority carriers in a semiconductor specimen. Derive any equations necessary for the calculation of these parameters and comment on the validity of any assumptions made.

A sample of germanium has dimensions 1 cm long (x-direction), 2·0 mm wide (y-direction) and 0·2 mm thick (z-direction). A voltage of 1·4 V is applied across the ends of the sample and a current of 10 mA is observed in the positive x-direction. A Hall voltage of 10 mV is observed in the y-direction when there is a magnetic field of 0·1 T in the z-direction.

Calculate (a) the Hall constant, (b) the sign of the charge carriers, (c) the magnitude of the carrier density, and (d) the drift mobility of the carriers.

(U.L. B.Sc. (Eng.) II Electrical, 1967)

6. A transistor grade sample of *n*-type germanium is found to have a resistivity of 1·5 and a Hall coefficient of magnitude $5{\cdot}4\times10^{3}$ cm^3/C. Determine the majority carrier density and mobility. What other parameter would be of importance in deciding whether this material was suitable for the fabrication of alloy junction transistors?

As the temperature of the sample is raised, it is found that initially the Hall coefficient remains constant while the resistivity increases somewhat, but at some higher temperature the magnitudes of both decrease sharply. Carefully explain this behaviour and suggest what observations would be made if the sample were cooled rather than heated.

(U.L. B.Sc. (Eng.) II, Electrical, 1967)

7. The Hall coefficient of a metal is found to be $1{\cdot}6\times10^{-4}$ cm^3 C^{-1}. What is the free electron concentration?

(Grad. Inst. P, II, 1966, part question)

8. Compare and contrast the operation of vacuum and semiconductor diode photocells. Why is silicon a particularly useful material to use in making solar batteries?

(Grad. Inst. P, II, 1966, part question)

9. Compare and explain the properties of intrinsic and extrinsic (impurity) semiconductors.

The electron concentration in intrinsic germanium is $2{\cdot}5\times10^{13}$ cm^{-3} and the electron and hole mobilities are 3800 cm^2 V^{-1} sec^{-1} and 1800 cm^2 V^{-1} sec^{-1} respectively. Calculate the electrical conductivity and the Hall coefficient.

(Grad. Inst. P, II, 1967)

10. Write accounts of two of the following: (a) phosphors, (b) ferroelectric crystals, (c) semiconducting compounds, (d) Hall effect devices.

(Grad. Inst. P, II, 1967)

CHAPTER 5

THE *p–n* JUNCTION

HAVING learnt a little of the behaviour of both *p*-type and *n*-type material in the previous chapters, the time has come to consider what occurs when a *p*-type semiconductor material forms a junction with *n*-type material.

5.1. SEMICONDUCTOR JUNCTIONS

It is very important for the student to realize that a semiconductor junction is quite different from the clear-cut boundary between (say) flat sheets of copper and glass. In a germanium *p–n* junction, the *p*-type material on one side is composed mainly of germanium, with a trace of *p*-type impurity; the *n*-type material on the other side is also mainly germanium, but with a trace of *n*-type impurity. The meeting of the two slightly different forms of germanium is a very intimate one, and the region in which it occurs is called the "*p–n* junction"; but in practice the junction may have been formed from one continuous germanium crystal (Fig. 5.1).

A useful analogy is to picture a very dilute salt solution carefully poured onto a very dilute sugar solution, to form a junc-

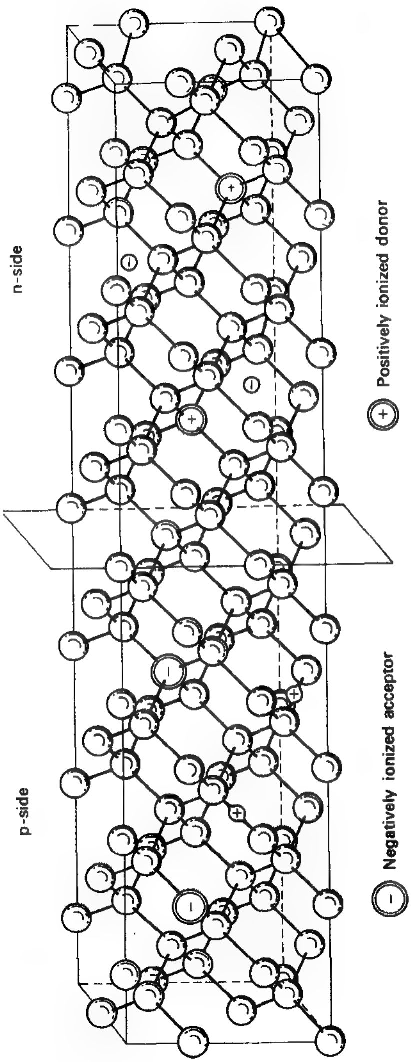


FIG. 5.1. The atomic structure of a *p–n* junction. (Courtesy of Alan Holden)

tion between the two solutions. Clearly the result is mainly water on either side of the junction. But after a while some salt molecules will have diffused into the sugar solution, and vice versa, due to the thermal motion of the salt and sugar molecules. The guiding principle is that the sugar molecules flow away from the regions of high sugar density to regions where it is less, and similarly for the salt molecules.

The freed electrons and holes in a semiconductor junction can wander through the lattice in a similar way, obeying the same rule.

5.2. DIFFUSION IN A SEMICONDUCTOR

Frenkel, a Russian physicist, was the first to appreciate the significance of diffusion in a semiconductor. Of course, charge carriers will drift if urged to do so by an applied electric field, but now we see there is a second cause of drift: a concentration gradient. An appreciation of both these causes is necessary for an understanding of a semiconductor junction.

Figure 5.2 illustrates the diffusion of charge carriers across the junction of *p*-type and *n*-type semiconductor material. The holes in the *p*-type experience a (relatively) high concentration of their own kind, and therefore diffuse away to the right and into the *n*-type region, leaving behind the negative acceptor ions. The opposite applies to the electrons in the *n*-type material. There is then a region in the crystal (on both sides of the junction) where it is depleted of its normal complement of charge carriers. This region, which can be as small as 10^{-4} m wide, is called the "depletion layer", and is a "space charge" region wherein the *p*-type side exhibits a net *negative* charge, and vice versa. (Remember that the crystal was originally electrically neutral.)

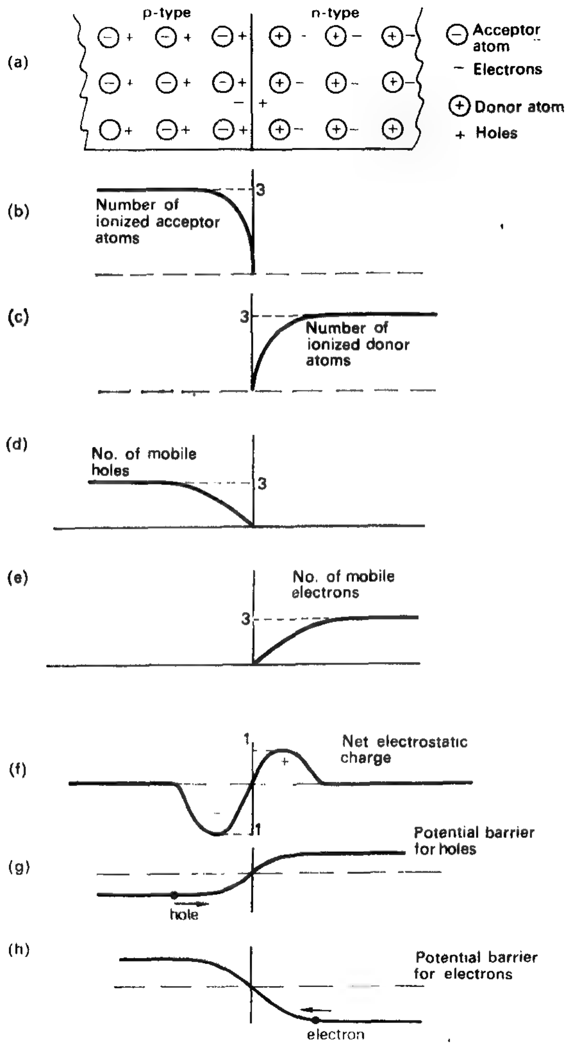


FIG. 5.2. The *p–n* junction

That these charges occur in the region of the junction is reasonably obvious, but if the diffusion process continued, the number of diffused electrons would eventually equal the number of diffused holes. However, this cannot occur, due to the finite "lifetime" of free electrons and holes. As soon as an electron enters the *p*-type material, the probability that it will recombine with a hole increases enormously. Again, the converse is true of holes entering the *n*-type material.

5.3. THE POTENTIAL BARRIER

It should be clear that the diffusion process just described does not continue indefinitely. Eventually the positive space-charge in the depletion layer of the *n*-type material (Fig. 5.2(f)) is strong enough to repel the approach of any more holes from the *p*-type material. The negative charge region similarly repels electrons, and equilibrium is established when this built-in potential difference across the depletion layer balances the diffusion process. This p.d. or *potential barrier* can be thought equivalent to a small cell across the junction, as in Figure 5.3. The magnitude of this p.d. is similar to the forbidden energy gap and is typically a fraction of a volt; but, of course, it has to be overcome before current can flow across the junction.

A useful way to imagine the equilibrium situation is to consider two equal and opposite currents of electrons, and two of holes. Figure 5.4 is really the same as Figure 5.2(h), and represents the potential barrier for the electrons. The electrons from the *n*-type region diffuse to the left giving a current I_D, whilst electrons thermally agitated out of the *p*-type region constitute an exactly compensating current I_G. The electrons constituting the current I_D have to "climb" the potential barrier to get from

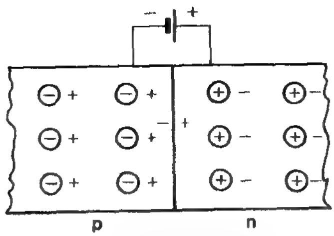


FIG. 5.3. The potential barrier inside a *p–n* junction behaves like a small cell

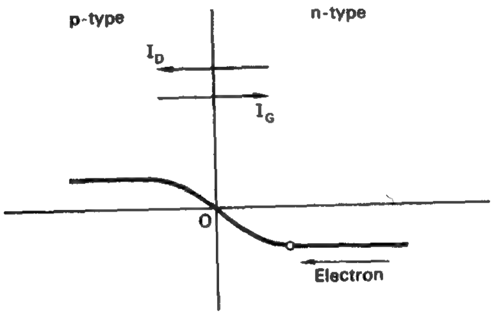


FIG. 5.4. The potential barrier for the electrons

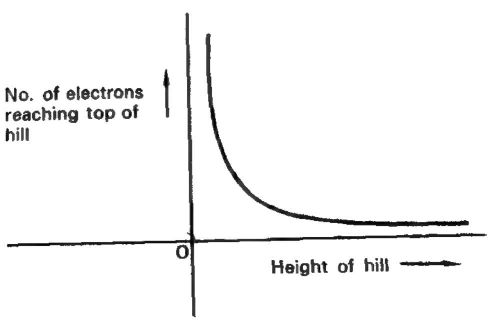


FIG. 5.5. A graphical representation of the electron statistics

the *n*-type to the *p*-type material until they get close enough to the barrier to "slide down" into the *n*-type side. Statistical analysis shows that the number of electrons which can get up the hill and thus contribute to I_D is related to the "height" of the hill, in the manner shown in Figure 5.5. We can see from this that as the height (potential barrier) decreases, the number of electrons able to climb it rises rapidly; therefore, so does I_D. But I_G is not dependent on the height of the "hill".

A similar picture can be used to explain Figure 5.2(g) for the holes.

The control of the height of the hill (and hence the width of the depletion layer) is of paramount importance, because it governs the value of I_D. We can achieve such control by an external source of p.d.

5.4. THE *p–n* JUNCTION WITH FORWARD AND REVERSE BIAS

If the external p.d. is applied positive to the *p*-side and negative to the *n*-side, the height of the "hill" is reduced because the potential of the external supply opposes the p.d. across the barrier; in addition, the electrons and holes experience the electric field applied by the external p.d., and acquire drift velocities (in opposite directions) superimposed on their random thermal motion. The electrons flow towards the positive pole of the external p.d., and the holes towards the negative. Again, we shall discuss only the electron flow, and assume that the hole flow about doubles the net current.

Figure 5.6(a) shows this effect; we say the *p–n* junction is "forward-biased". Figure 5.6(b) shows an *increase* in the height of the "hill", because the external p.d. has been reversed. The junction is now said to be "reverse-biased".

Notice in Figures 5.4 and 5.6 that I_G, the electron current thermally generated from the *p*-side, has a magnitude independent of the size of the potential barrier; but as the height of this

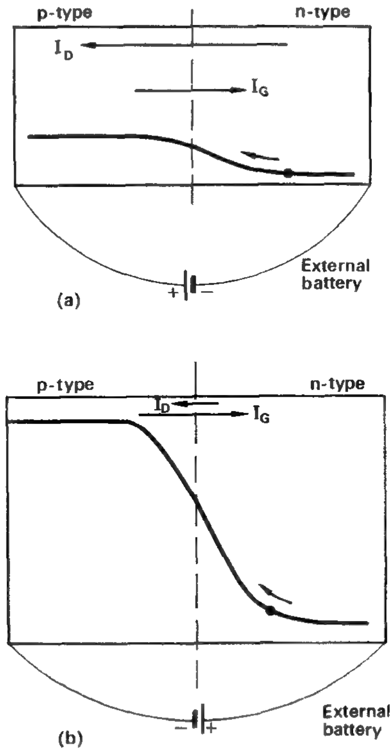


Fig. 5.6. (a) A *p–n* junction with forward bias, (b) a *p–n* junction with reverse bias

hill is continuously reduced, I_D, the electron current diffusing from the *n*-type to the *p*-type region, increases. When I_D becomes greater than I_G (in forward bias), considerable currents can flow across the junction; but when I_D is less than I_G (in reverse bias),

only very little current can flow. (Obviously with no bias, $I_D = I_G$, and the net current is zero.)

So the current/voltage characteristic of a p–n junction is of the form shown in Figure 5.7. The "saturation" (maximum) current in the reverse bias situation is just the amount I_G. The characteristic is blatantly non-ohmic, and the fundamental property of a biased p–n junction is that of rectification. This may involve currents of a few microamps, for example in radio wave detec-

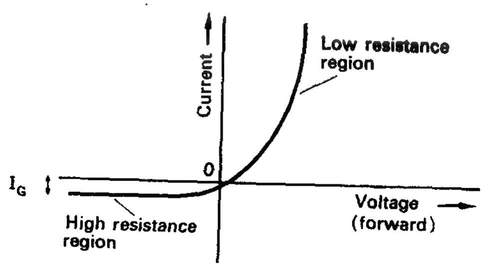


FIG. 5.7. The current/voltage characteristic of a p–n junction

tion, or as high as hundreds of amperes, for example in electroplating and cathodic protection of the metal hulls of ships. The p–n junction also has an important role in the logic circuitry of computers.

The reader may find it helpful to consider the effects of bias in terms of the energy-band diagrams (Fig. 3.4). Carrier diffusion occurs across the p–n junction until the p.d. created is large enough to prevent further diffusion; then the Fermi levels E_F on each side are in line (Fig. 5.8(a)). If the junction is forward-biased (Fig. 5.8(b)) the potential barrier is decreased; then the electrons in the conduction band on the n-type side can easily get into the conduction band on the p-type side, and the holes in the valence band on the p-type side can easily get into the valence band on the n-type side. The carriers which cross, of

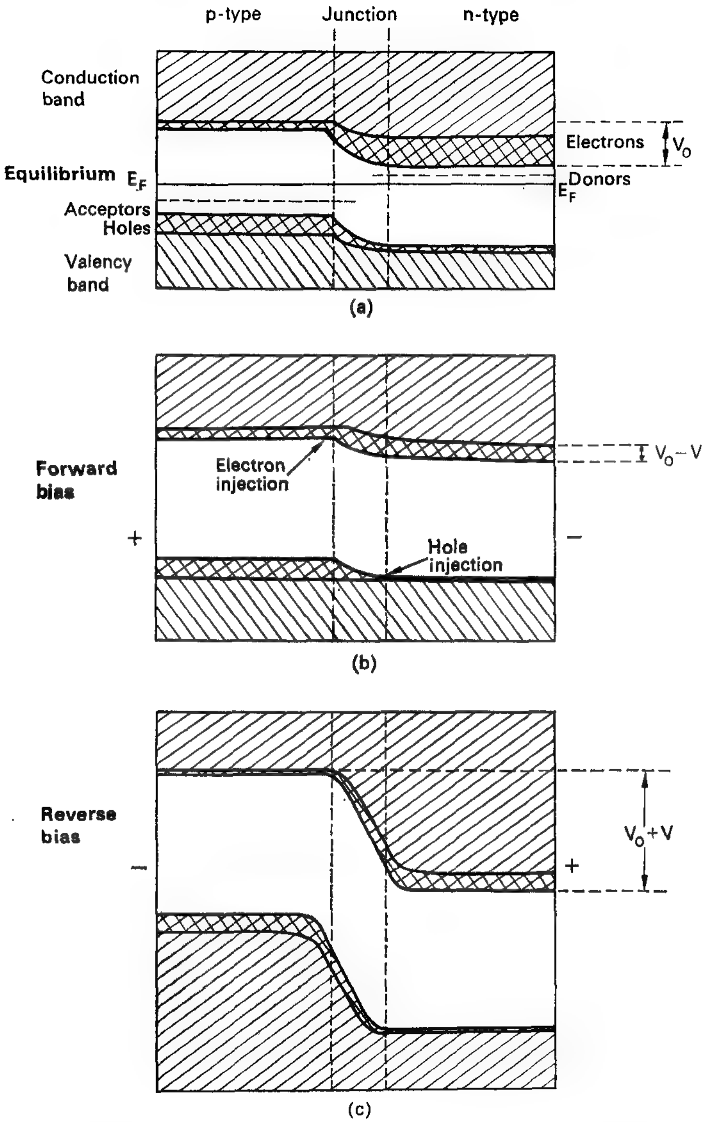


FIG. 5.8. (a) The energy band diagram of a *p–n* junction in equilibrium, (b) the energy band diagram of a *p–n* junction with forward bias, (c) the energy band diagram of a *p–n* junction with reverse bias

course, are "in foreign territory", and only survive a short time before recombining with an opposite. When the junction is reverse-biased (Fig. 5.8(c)) the potential barrier is increased, making it fairly certain that electrons in the n-type conduction band will stay there, as will the holes in the valence band of the p-type material. But any electrons that *may* be present in the conduction band of the p-type material (and holes in the valence band of the n-type), on reaching the hill in the course of random thermal motion, will easily slide down, constituting a slight current. (See Expt. 5.1.)

5.5. THE RESISTANCE OF A p–n JUNCTION

The characteristic in Figure 5.7 obeys the equation

$$I = I_G(e^{qV/kT} - 1) \tag{5.1}$$

where I_G is the saturation current in the reverse direction,
I is the current when the applied voltage is V,
q is the charge associated with each carrier,
T is the absolute temperature, and
k is Boltzmann's constant.

A useful expression can be derived for the approximate resistance in the forward direction, if we differentiate eqn. (5.1) with respect to voltage:

$$\frac{dI}{dV} = I_G \cdot \frac{q}{kT} \cdot e^{qV/kT} = \frac{1}{r} \tag{5.2}$$

where r = small-signal A.C. resistance in the forward direction (i.e. ratio of small voltage change to small current change when the bias is at all stages forward).

But if V is greater than about 100 mV,

$$I \simeq I_G e^{qV/kT} \tag{5.3}$$

(because the I_G term in (5.1) may be neglected in comparison).

Substituting for I from (5.3) into (5.2),

$$\frac{1}{r} \simeq \frac{q}{kT} \cdot I$$

and kT/q has the value of about 25 mV when $T = 300°K$ (about room temperature), so

$$\frac{1}{r} = \frac{I}{25} \text{ (where } I \text{ is in mA),}$$

and $$r \simeq \frac{25}{I}.$$

The student will realize that if V has a large negative value, substitution of this in eqn. (5.2) shows r to be very large indeed. But an interesting phenomenon occurs if V is made sufficiently negative: a "breakdown" results, and a large current then flows in the reverse direction. This so-called "Zener" effect is treated in Section 6.3.

In Chapter 6 we proceed to describe how individual p–n devices are made, and their uses in addition to rectification.

EXPERIMENT 5.1: Characteristics of a p–n diode.

Apparatus required:

General purpose low-current junction diode, e.g. 1GP5 (Radiospares),
Transformer giving about 5 V output,
Resistor, 1 kΩ, 1/4 W,
Oscilloscope with X input facility (the Advance Serviscope Minor is ideal),
Connecting leads.

Procedure. Set up the circuit of Figure 5.9. If using the Serviscope Minor, the Y gain will need to be set to about 5. Otherwise adjust X and Y gains, and shift controls, until a trace similar to Figure 5.10 is seen. As shown in the diagram, the current (y) axis is necessarily reversed.

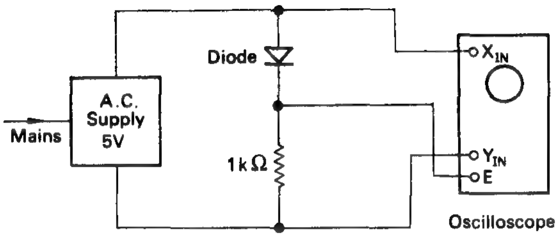


FIG. 5.9.

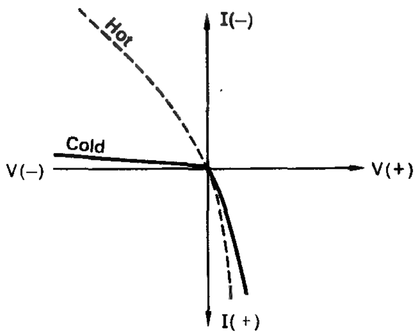


FIG. 5.10.

Heat the diode for a second or two, but no more, with a match. The trace alters towards the dotted position. This indicates much higher reverse current, and a lower reverse resistance. As soon as the diode is allowed to cool, the normal trace reappears.

QUESTIONS

1. Explain briefly the following terms: depletion layer, space charge region, "lifetime" of free electrons, "hole", "potential barrier".

2. Find the forward and reverse slope resistances of a given p–n junction diode, from the following data:

I (mA)	−0·5	−0·3	1·0	10·0
V (volts)	−2·0	−0·3	0·3	0·5

Sketch the graph, indicating what you would expect to happen if the reverse voltage were increased gradually to a very high value.

3. How are the forward and reverse slope resistances of a given *p–n* junction diode affected by heating? Give orders of magnitude, and outline the reasons.

4. Sketch the V/I characteristic for a junction diode and indicate (a) how it is affected by temperature, and (b) how it differs from the theoretical relationship.

The reverse saturation current I_s of a germanium junction diode is 15 μA at a temperature of 20°C and increases by 11% per °C. Calculate the forward voltage drop across the diode with a current of 100 mA at a junction temperature of 55°C. Mention any assumptions made in the calculation.

Boltzmann's constant $k = 1{\cdot}38 \times 10^{-23}$ J K^{-1}
Electron charge $e = 1{\cdot}6 \times 10^{-19}$ C

(I.E.E., Part 3, Dec. 1967)

5. Describe by means of energy level diagrams, or otherwise, how rectification can occur at a metal–semiconductor (*n*-type) junction. The work functions of the metal and semiconductor are respectively ϕ_m and ϕ_s with ϕ_m greater than ϕ_s. What explanation can be given of the fact that in practice the rectifying characteristic is not a function of the contact potential, but is largely independent of ϕ_m?

(Grad. Inst. P, Part 2, 1967)

6. Derive the equation relating direct current and voltage for a *p–n* abrupt-junction diode with a uniformly doped base. Assume the width of the base is small compared with the diffusion length of minority carriers in it. Indicate clearly the significance of the approximations made. In the circuit shown in Figure 5.11 the alloy-type germanium diode is initially

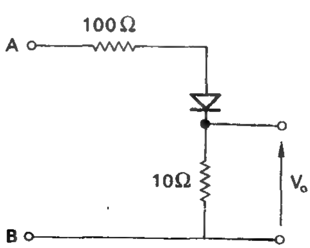


FIG. 5.11.

passing a forward current of 1 mA. At time $t = 0$, the potential at A is changed to -10 V. Draw and explain the shape of the output voltage waveform, V_0.

(Grad. Inst. P, Part 2, 1967)

7. A silicon junction diode has a reverse saturation current of 1 μA at 27°C. What is its highest operating temperature if the maximum acceptable reverse current is 1 mA?

(Energy gap of silicon is 1·1 eV; kT/e at 27°C is 0·025 V.)

(Grad. Inst. P, Part 2, 1966, part question)

8. Explain the phenomenon of depletion capacitance at a *p–n* junction, and show how this capacitance depends on the applied voltage for an abrupt junction.

If the junction is graded such that the impurity concentration in the *p*- and *n*-regions is proportional to the distance from the centre of the junction, calculate the new dependence of depletion capacitance on applied voltage.

9. Sketch the electron energy level diagram for the contact between a metal, work-function ϕ_m, and an *n*-type semiconductor, work function ϕ_s, (a) for zero bias, (b) for forward bias and (c) for reverse bias, when ϕ_m is greater than ϕ_s. By reference to this diagram, or otherwise, explain the rectifying action of such a contact. What type of contact is produced if ϕ_m is less than ϕ_s, and why?

Why are metal–semiconductor diodes frequently used in preference to *p–n* junction diodes when very high-speed operation is required?

(U.L. B.Sc. (Eng.), Part 2, Electrical, 1967)

10. Write a brief explanation of the rectifying property of a *p–n* junction. Obtain an expression, in terms of the properties of the semiconductor, for the reverse saturation current of the junction, and discuss any assumptions made.

(C. of E.I., Electronics, Dec. 1966)

CHAPTER 6

p–n JUNCTION DEVICES

6.1. PRACTICAL JUNCTIONS

It was explained in Section 5.1 that a *p–n* junction is formed from one continuous crystal of a suitable material, germanium or silicon (say), with opposite sides of the junction being slightly and differently doped with suitable impurities. Initially *grown junctions* were produced, by adding the right impurities to the single crystal growth as it was pulled from the melt: this process is rarely used now and we shall consider three modern processes for producing *p–n* junctions.

6.1.1. The alloy junction

This method was a popular one in the 1950s, typical examples being the alloying of aluminium into silicon and indium into germanium to form *pnp* transistors. In the case of the latter, germanium melts at 950°C but indium melts at 156°C: so a small pellet of indium (an acceptor impurity) is placed on a wafer of *n*-type germanium and the pair heated. At a temperature of 156°C the indium melts (see Fig. 6.1) and as the temperature is increased to about 600°C, the molten indium dissolves some of the germanium and the solid–liquid interface moves down into

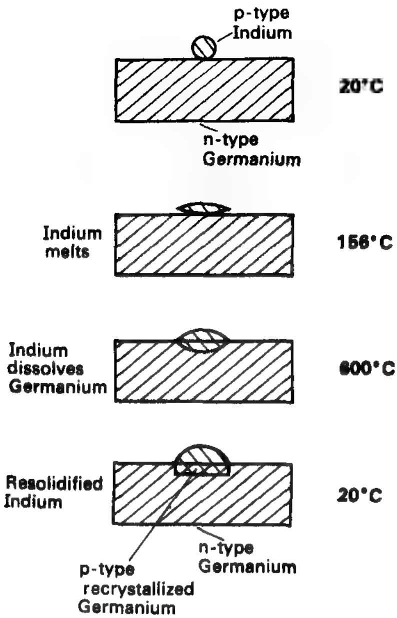


FIG. 6.1. The alloy junction process

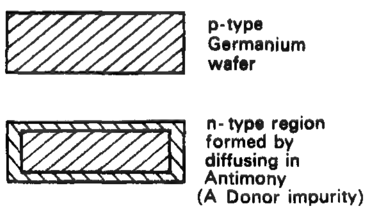


FIG. 6.2. The diffused junction process

the germanium to a depth determined by the temperature and volume of the indium. When the temperature is reduced to room temperature, the germanium recrystallizes out of the indium–germanium solution back onto the germanium wafer, with the

same crystalline orientation: but this germanium is doped with indium, and this recrystallized layer is *p*-type, so we now have a *p–n* junction.

6.1.2. The diffused junction

In the diffusion process a liquid phase is not used (cf. the alloy junction method); instead, the impurity atoms are present as a gaseous atmosphere around the solid semiconductor wafer, and diffuse directly into it. The depth of impurity concentration can be controlled to about 10^{-7} m, and this is of considerable importance in the production of very thin base regions for transistors with good high frequency performance.

A *p*-type germanium wafer, for example, is heated to about 100°C below its melting point in a gaseous atmosphere containing antimony (a donor impurity). A small proportion of the antimony atoms diffuse into the germanium, introducing an excess of donors and turning the original *p*-type layer into an *n*-type one. So just inside the germanium wafer a *p–n* junction is formed (see Fig. 6.2). It is an interesting but not fully understood fact that, whereas the above process takes place at about 10^{-6} m per hour, diffusing acceptor impurities (e.g. gallium) into germanium is so slow a process as to make it almost impracticable.

Silicon can be used instead of germanium, of course, suitable impurities being boron (acceptor) and phosphorus (donor).

6.1.3. Epitaxial junctions

A very thin layer of suitable semiconductor is grown on a wafer of the same material in the epitaxial process, the same crystalline structure growing from the wafer (called the "substrate") into the deposited layer. As the same geometrical axis continues through the substrate and the new layer, we use the word "epitaxial".

If we begin with a *p*-type germanium wafer as the substrate, then a compound such as germanium tetrachloride mixed with hydrogen is passed over a donor impurity and fed into a chamber containing the substrate, and the temperature raised to about 800°C. A chemical reaction occurs with the $GeCl_4$ dissociating, the hydrogen combining with the chlorine, and the *n*-type germanium depositing on the *p*-type substrate as an epitaxial layer. Thus a *p–n* junction is formed. The process can work in reverse, of course, starting with an *n*-type substrate. Alternatively silicon may be used similarly, but a temperature of 1100–1200°C is needed. The rate of growth of an epitaxial layer is about 10^{-6} m per *minute*.

Finally we will just note that high-resistivity *n*-type layer (for example) can be grown on a low-resistivity *n*-type substrate: this apparently odd combination is useful in some devices, for example "heterojunctions" (Section 10.5).

We shall return to junction manufacture in Section 8.1, when we discuss briefly the technology of transistor manufacture.

6.2. RECTIFIERS

We have seen in Chapter 5 that a *p–n* junction presents a low resistance if forward biased, and a very high resistance when reverse biased. So the *p–n* junction makes an efficient *rectifier* (i.e. it can turn A.C. into D.C.); it is called a *junction rectifier*. Figure 6.3 illustrates the rectification principle (see Expt. 6.1). The alternating voltage input is applied in series with the load resistor and rectifier in series. When the top terminal of the input goes positive, it biases the junction in its forward, or low resistance, direction, and a large current passes through the

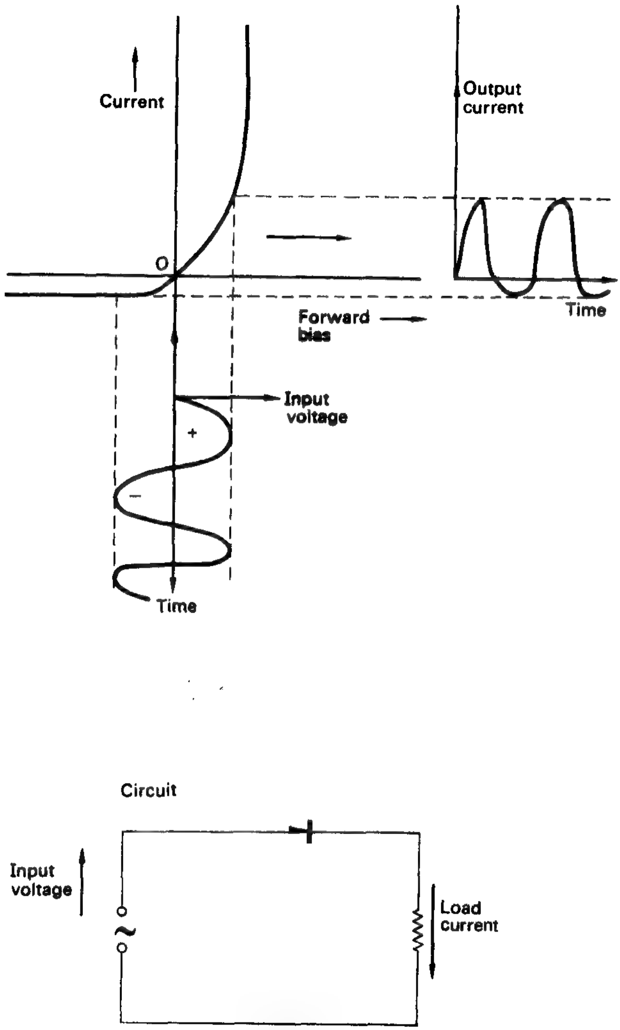


FIG. 6.3. The rectification principle and circuit

load. However, when the top terminal of the input goes negative, it biases the junction in its reverse (very high resistance) direction, and only a negligible current flows through the load. A perfect *p–n* junction would not conduct at all in the reverse direction; the rectifier designer minimizes the reverse current by using a semiconductor material with a wide forbidden energy gap and a high concentration of impurities, so reducing the number of intrinsic carriers (which are the source of reverse current). But this unfortunately also results in a reduction of the reverse voltage that the junction can withstand before the onset of "avalanche" (see Section 6.3). Silicon has a wider forbidden energy gap (1·21 eV) than germanium (0·75 eV), thus keeping the reverse leakage current very small and enabling the junction to work at a higher temperature. A rectifier must be manufactured (by alloying or diffusing) so as to have low-resistance connections to the *p* and *n* sides of the junction, and low thermal resistance between the junction and some external heat dissipating surface, because the heat generated by the internal power loss must be carried away or the junction may melt. The internal power loss is the product of the forward voltage drop and the forward current, divided by two because it occurs for only half the time of each cycle of the alternating supply. With silicon the power loss due to reverse current is negligible. Silicon junction rectifiers of area several cm^2 are now produced, able to handle about 300 A with reverse voltages of up to 1500. They are ideal for high D.C. requirements, e.g. industrial electroplating, and cathodic protection of ships' hulls. Such rectifiers are clamped to a massive metal fin system which dissipates the heat, and is called the "heat sink".

Medium power junction rectifiers are often "stud mounted", bolted directly onto the metal chassis of the circuit they serve; whereas low power ones (less than 1 A) are wire ended. The latter are finding increasing use in domestic electronics, replacing the

thermionic diode in an otherwise identical rectifying and smoothing circuit (see Vol. 1).

It is worth noting that the silicon junction rectifier has superseded the older copper/copper oxide and selenium–tin/cadmium rectifiers in most applications; the limitations of the latter are about 100 mA cm^{-2} and 50°C.

6.3. DIODES

Although the *p–n* junction rectifier just described is, strictly, a diode, in semiconductor terminology it is referred to as a rectifier, and the word "diode" is reserved for those *p–n* junctions used for such general circuit functions as detection of radio signals, limiting, mixing, switching, etc.

6.3.1. The point-contact diode

This device is the "crystal and cat's whisker" used in the early days of radio reception, and from which we get the term "crystal set". Basically it is a sharp-pointed tungsten wire pressed against a germanium crystal (though other materials have been used successfully) (see Fig. 6.4). The working of this device was not understood in its day, and was rendered obsolete as a radio-wave detector by Fleming's thermionic diode. However, during the development of radar in the Second World War, the need to use microwave frequencies (wavelengths of 3–10 cm)

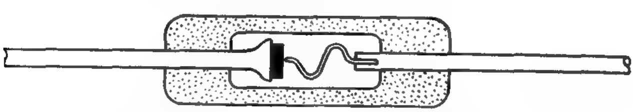


FIG. 6.4. A germanium point-contact diode. (Courtesy of R. G. Hibberd)

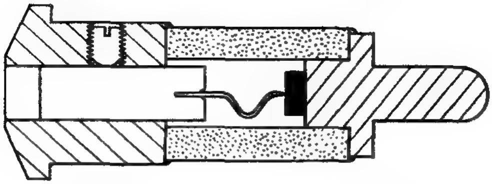


FIG. 6.5. A silicon cartridge microwave diode. (Courtesy of R. G. Hibberd)

saw the rebirth of interest in semiconductor diodes, because of the frequency limitations of thermionic diodes. The silicon point-contact diode (see Fig. 6.5) was useful as a detector and frequency-changer at frequencies up to 30 GHz (3×10^{10} c sec^{-1}), and the germanium one up to about 100 MHz, but both are now being superseded by high-frequency *p–n* junction diodes.

6.3.2. The gold-bonded diode

Basically this is an *n*-type germanium point-contact diode with the tungsten wire replaced by a gold wire, about 1/1000 inch in diameter, to which has been added a small percentage of an acceptor impurity such as gallium or indium. A pulse of current is used in manufacture to fuse the gold to the germanium, effectively producing a *p–n* junction. This device was found useful in computer circuitry because of its very small forward voltage drop and leakage current; but again it has been superseded, in this case by *p–n* junctions of small area.

6.3.3. The junction diode

These are similar to the rectifiers of Section 6.2 in that they are *p–n* junction diodes, but in this section we specifically mean low-power types. A typical sub-miniature silicon diffused junction diode is shown in Figure 6.6. The junction area is very important:

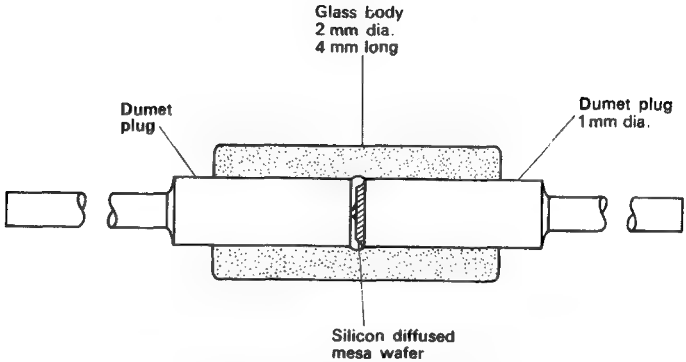


FIG. 6.6. A sub-miniature silicon-diffused junction diode.
(Courtesy of R. G. Hibberd)

TABLE 6.1

	Germanium	*Silicon*
Reverse voltage rating	≏ 700 V	Highest ≏ 2000 V (so useful for high voltage rectifiers)
Forward voltage drop	≏ 50 mV Lowest (so gives highest efficiency in low voltage rectifiers)	≏ 700 mV
Junction temperature rating	≏ 90°C	≏ 200°C (again, more suitable for high power rectifiers)
Reverse current		Lowest
High frequency performance	Best	

for a general purpose diode a junction diameter of 0·15 mm is typical, the forward current being about 200 mA for a forward voltage drop of 1 V. The leakage or reverse current can be as low as 10 nA (1 nA = 10^{-9} A) up to a reverse voltage of about 300 V. For this junction area the junction capacitance is about 8 pF at zero bias. If the junction diameter were halved, the capacitance would fall to about 2 pF, making the diode suitable for high-speed switching, with switching times of a few nanosec. ("Switching time" in this context is the time it takes the diode to reduce a current of 10 mA to zero, as its bias is varied.)

The *p–n* junction diode is a most important device and a great variety are manufactured, from those working at a few milliamps to the rectifier types carrying hundreds of amperes. As an example, a designer may be called upon for a diode with low capacitance, low forward resistance, high forward current, high reverse resistance, low slope in the breakdown region, or high breakdown voltage; and it might be produced in germanium or silicon. Some of the contrasting properties of germanium and silicon diodes are shown in Table 6.1.

6.4. COMPARISON OF *p–n* JUNCTION DIODES WITH THERMIONIC VALVE DIODES

As a *p–n* junction diode has no heater that can burn out, no oxide cathode coating that can decompose and lose emission, and no glass envelope to "soften" or break, longer life can be expected if the ratings are kept to. The junction diode is light and needs no special holder; it has a low capacitance, and it cannot introduce "hum" into a circuit. Unfortunately it has an appreciable reverse current, whereas a thermionic diode does not have this defect. Its electrical characteristics are far more

temperature-dependent than those of the thermionic counterpart, and this is the commonest reason for failure; but it is easily allowed for in design.

6.5. THE VARACTOR DIODE

We saw in Chapter 5 that a *p–n* junction involves two layers of electric charge separated by a short distance—in other words, it has the basic feature of a capacitor. In a parallel-plate capacitor the capacitance depends on the area of the plates, the medium between them, and its thickness. In any given *p–n* junction the effective distance between the two opposite charge concentrations depends on the voltage across the device—in other words, we have a capacitor whose value is a function of the voltage across it. Quantitatively, the junction capacitance varies inversely with the reverse-bias voltage.

Diodes made specifically to utilize this effect are called "varactors", and have an extremely small junction area. A typical silicon-diffused varactor might have a junction diameter of 0·03 mm and exhibit a capacitance of 1 pF and a forward resistance of 3 Ω at zero bias, giving it a cut-off frequency of about 50 GHz. Gallium arsenide varactors are now available with a cut-off frequency of some 200 GHz.

The device is used for frequency multiplication in transmitter circuits for ultra-high frequencies (for example, 40 watts of power at 144 MHz can be used with a varactor and tuning components only to give about 20 watts output at 432 MHz, and the component is easily capable of dissipating the lost 20 watts as heat), and as an electrically controlled tuning capacitor in radio receivers.

6.6. THE ZENER DIODE

In a reverse biased *p–n* junction the electric field intensity (or potential gradient) is very high, because although the p.d. across the junction may only be a few volts, the distance over which it is applied is only a few atom diameters(!). Such a high

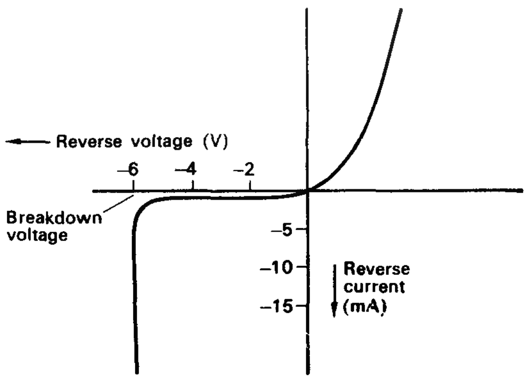


FIG. 6.7. Typical Zener diode characteristic

field can accelerate electrons in the junction region, which came from the conduction band on the *p* side, to such high energies that they can knock electrons from atoms in the junction region. Electrons so released are also accelerated, and knock off more electrons. This "avalanche" or "chain reaction" rapidly grows to an appreciable current, with very little increase in reverse voltage. Diodes made to use this effect are called Zener diodes, after the physicist who first explained the mechanism of the avalanche. A typical Zener characteristic is shown in Figure 6.7. Silicon shows a more sharply defined "knee" than germanium, and this is further enhanced by manufacturing from low resistiv-

ity material. The breakdown voltage can be made to be anything from about 3 to 300 V. The device finds considerable use as a voltage limiter or stabilizer (cf. gas-filled diode voltage stabilization, Vol. 1), and as a voltage reference. Gas-filled diodes cannot be used in such ways below about 30 V, because of the higher ionization potential of gases.

A detailed analysis of the breakdown shows that two effects are in fact involved, called Zener tunnelling and avalanche breakdown. It is not always possible to distinguish the two in any practical case. The former occurs at a lower voltage. Specifically *avalanche* diodes find use in electrostatic dust precipitation plants where voltages of say 50 kV are involved.

6.7. PHOTOCELLS

We saw in Section 4.4 that a basic property of semiconductors is photoconductivity, i.e. the resistance depends on the intensity of incident light. A *p–n* junction is also light sensitive, as might be expected, and is called a photoconductive photodiode. The principle must be explained with reference to the energy band diagram for a *p–n* junction (Fig. 6.8). The *p–n* junction is reverse biased, so any free charge carriers generated in the junction region are swept across the junction, the electrons "rolling down" the conduction band attracted by the positive charge, and the holes "climbing up" the valency band attracted by the negative charge. This normal reverse (or leakage) current can be detected with a microammeter. When light energy enters the junction layer, hole–electron pairs are created, thereby increasing this leakage current. Thus a *p–n* junction, reverse-biased, makes a useful photocell; infra-red sensitive versions are made from germanium and indium antimonide, and are preferable to the photo-

conductive cell described in Section 4.4 because in the latter the energy gap is small and the concentration of intrinsic holes and electrons is appreciable at room temperature, making the change in current due to incident light difficult to measure. The much smaller reverse current of a photodiode in the dark (referred to as the *dark current*) makes for a large percentage change in current with light intensity.

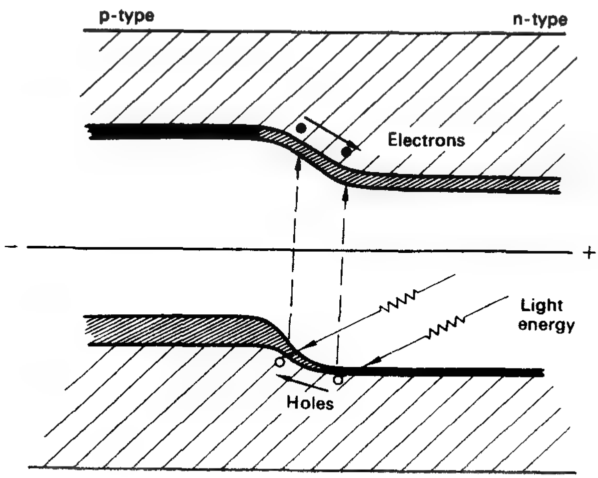


FIG. 6.8. Photoconduction at a *p–n* junction

There are photoconductive devices for use with reverse voltages up to 100 V; for instance, a germanium type has a sensitivity of up to 30 mA per lumen deep in the infra-red region.

It is not essential to apply an external p.d. to the *p–n* junction in this application, as the natural potential "stop" may be sufficient to sweep the photo-generated charge carriers out of the junction region, and to send them round an external circuit containing a sensitive galvanometer. In this application the device is termed *photovoltaic*, and is commonly known as a

solar cell. Silicon solar cells can produce about 60 W/m², with an efficiency (light energy to electrical energy) of about 12% (compare this with the 0·6% efficiency of the selenium photovoltaic cell of Vol. 1, Sect. 11.2.1). They have been used for some time as the source of electrical energy in satellites, as photographic exposure meters and controllers, and in alarm circuits (see Expt. 6.2). The construction of a typical solar cell is shown

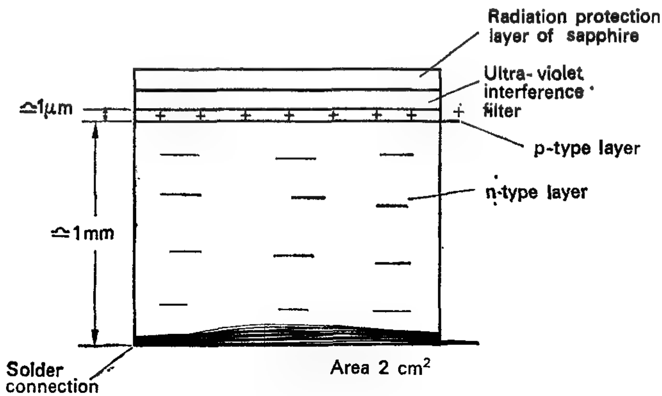


FIG. 6.9. Construction of a typical solar cell

in Figure 6.9. It should be made clear here that if a junction diode is required to be insensitive to light changes, it must be enclosed in a light-tight case.

We proceed in Chapter 7 to discuss a *p–n* junction device—the transistor—that is of exceptional importance in its own right.

EXPERIMENT 6.1: The *p–n* junction as a rectifier.

Set up the circuit of Figure 6.10. The REC51A (Radiospares) is capable of rectifying half an ampere at 250 V, and is not overrun in this experiment. (If a "Serviscope Minor" is used, the gain should be about 2, and the time-base on range 2.)

With *C* disconnected, the bulb glows dimly and the flicker is conspicuous. Average meter readings are 0·2 A and 3 V output, but moving-coil meters cannot be regarded as accurate on rectified unsmoothed current.

When *C* is connected the bulb becomes much brighter and the flicker ceases; meter readings are now typically 11 V and 0·4 A. The oscilloscope shows the great difference between unsmoothed half-wave rectified current, and the same smoothed. Further smoothing may be introduced, and its

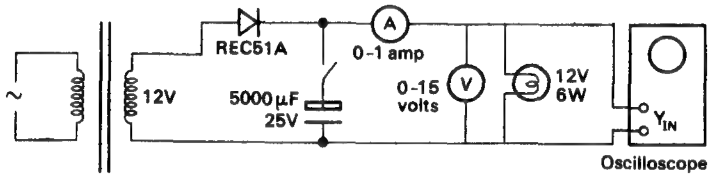


FIG. 6.10.

effect observed, along the lines of Vol. 1, Expt. 5.2; but larger capacitors, and high-current chokes, will be needed.

EXPERIMENT 6.3: Zener diode for voltage stabilization.

Apparatus required:

Zener diode rates at about 10 V, 1·5 W (e.g. Mullard BZY96C10).
Power *p–n–p* transistor, e.g. OC26.
Resistors, 6·8 Ω 15 W and 68 Ω 2 W (or near).
Variable resistances, 10 Ω 15 W and 100 Ω 2 W (or near).
Voltmeter, 0 – 20 V f.s.d.
Ammeter, 0 – 2 A f.s.d., and milliammeter, 0 – 200 mA f.s.d.
D.C. supply, fully variable, 6–20 V, up to 1·5 amp.

Procedure. (a) Set up the circuit of Figure 6.11(a). For various values of supply voltage *greater than* the Zener voltage, plot graphs of load current against load voltage.

(b) The circuit of Figure 6.11(b) shows how a power transistor may be used with a Zener diode to act like a Zener diode of higher power-handling capability. Again, for various values of supply voltage greater than the zener voltage, plot graphs of load current against load voltage.

The Zener diode, or Zener diode and transistor, draw sufficient current through the series resistance to maintain the desired voltage. Note that the higher the supply voltage, the greater the range of load current over which voltage stabilization works; but the greater also the power lost as heat in the series resistor, diode, and transistor.

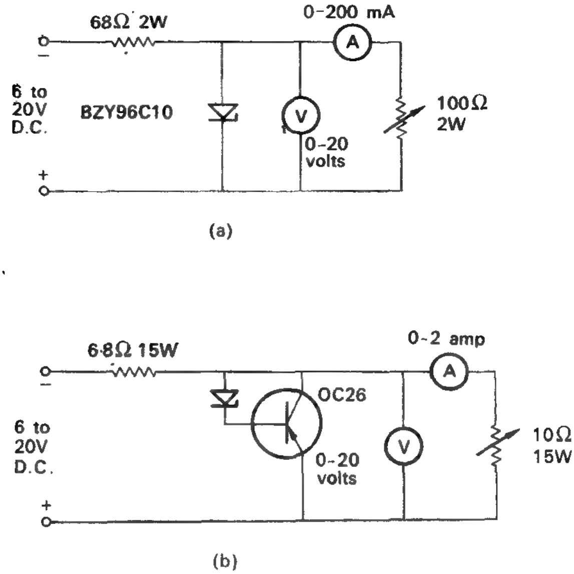


FIG. 6.11.

QUESTIONS

1. Compare the advantages in manufacture and use of alloy, diffused, and epitaxial junctions.

2. What are the properties required of a varactor diode? How are they achieved? What are the uses of the device? Compare these features with the corresponding ones for a Zener diode.

3. Mention three advantages of semiconductor voltage regulator diodes as compared with gas-filled regulators in their application to shunt stabilizer circuits.

The circuit diagram of a simple D.C. shunt voltage stabilizer is given in Figure 6.12(a), together with the Zener diode reverse characteristic, Figure 6.12(b). Assuming that the maximum permissible Zener diode current is 850 mA, calculate the range of variation of input voltage V_i over which

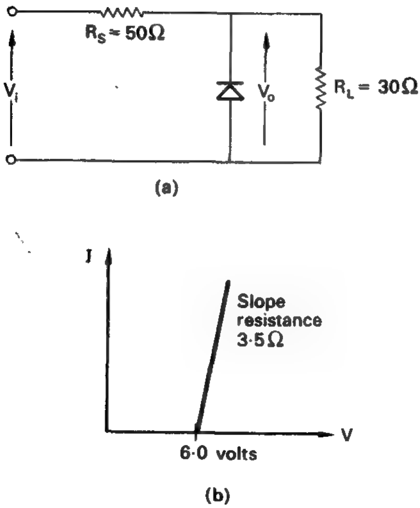


FIG. 6.12.

stabilization may be achieved. Find also the corresponding range in the value of the output voltage V_0. If the input voltage is maintained at its maximum permissible limit, calculate the percentage reduction in output voltage when R_L is decreased to 20 Ω.

(U.L.C.I, Industrial Electronics I, May 1967)

4. Give a sketch showing the components of a selenium rectifier unit and give an account of its action. Sketch the voltage–current characteristic for the forward and reverse voltages.

With the aid of a diagram show how these rectifiers can be connected in bridge form to give a full-wave rectification. Indicate on the diagram the direction of current flow for each half-cycle.

(U.L.C.I., Electric Technology, June 1966)

5. A battery charger is arranged as shown in Figure 6.13. The secondary winding of the mains transformer produces a sinusoidal waveform having a peak voltage of 24 V. The total series resistance in the circuit of the secondary winding is $R_s = 3\ \Omega$ and the forward voltage drop across the semiconductor diode is negligible. Derive an expression for the average current supplied to the battery. Calculate the value of this current and the

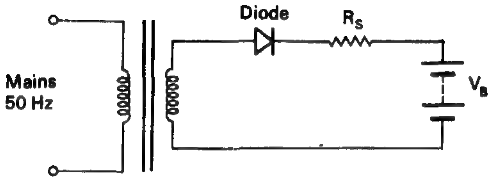


FIG. 6.13.

average power dissipated in the battery, the voltage of which during charging is $V_B = 12$ V.

(U.L. B.Sc. (Eng.), Part I, 1967)

6. Write brief notes on Tunnel and Zener diodes.

7. In Figure 6.14 all four diodes have similar characteristics, specified by $I = T_s\ [\exp(eV/kT)-1]$, where $I_s = 10\ \mu$A and $T = 300°$K. R is a $100\,\Omega$ resistor. Find the approximate voltages across D_4 and R when A is 1 V positive with respect to B. Discuss the factors which affect the reverse saturation current of a semiconductor p–n junction diode.

(Grad. Inst. P, Part 2, 1966)

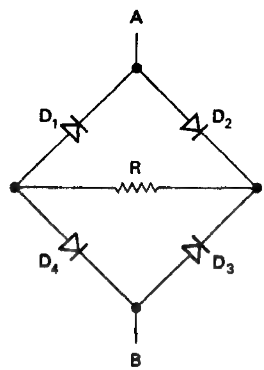


FIG. 6.14.

8. Give a qualitative account of the avalanche and Zener breakdown mechanisms in semiconductor p–n junctions. Draw and describe the operation of a simple solid-state stabilized power supply incorporating a reference diode.

(Grad. Inst. P, Part 2, 1966)

9. In the circuit of Figure 6.15 the capacitors have no loss, and the diodes have zero resistance in the forward direction and infinite resistance in the reverse direction. Initially C_1 and C_2 are uncharged and an input voltage consisting of a uniform series of positive-going rectangular pulsés of magnitude V_i is then applied. Draw the corresponding output voltage waveform and develop an expression relating it to the input voltage. Suggest a practical application of this circuit using a sinewave input.

(Grad. Inst. P, Part 2, 1967)

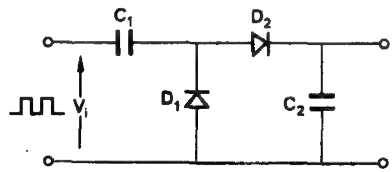


FIG. 6.15.

10. The theoretical V/I characteristic of a uniform plane junction diode at 300°K may be shown to have the approximate form

$$I = -I_s\,(e^{39V}-1)$$

where I_s is the saturation current.

Sketch the theoretical and practical V/I characteristics for a typical diode and explain the difference between the curves.

When 1 V reverse bias is applied to a diode the current is 0·1 μA. With 1 A flowing in the forward direction the potential drop across the diode terminals is 1·2 V. Determine the bulk resistance of the diode and hence calculate the power dissipation for a current of 2 A.

(I.E.E. Part 3, June 1967)

CHAPTER 7

THE TRANSISTOR

IN 1947 at the Bell Telephone Laboratories in the U.S.A., Bardeen, Brattain, and Shockley were investigating the surface potentials around a point contact (between a needle and a piece of semiconductor), by using a second needle to probe the electric field. They found that when the separation of the two points was about 1/1000 in., a change of current at one contact influenced the current in the second contact. This effect constitutes *current amplification*, and its significance was immediately appreciated: a change in current in a *low* resistance part of the device (the emitter-base junction) caused a similar change of current in a high-resistance part (the base-collector junction), thus constituting power (and voltage) amplification. This "transfer of resistance" may have been the origin of the name given to the device so formed: the *trans*fer res*istor*, or transistor.

In comparison with the vacuum triode the transistor is smaller, lighter, uses less power, and needs no heater supply. However, the "point contact" transistor described above had some serious limitations, i.e. the extreme difficulty of manufacturing devices with similar characteristics, the high cost of manufacture, the fragile nature, and the high noise factor. But the possibility of "solid-state amplification" was proved beyond all doubt, and encouraged an intensive research programme which soon pro-

duced the *junction* transistor. In this, instead of two point contacts on a semiconductor slice, two *p–n* junctions are formed inside the crystal. The junction transistor has superseded the point-contact type, and has given rise to a whole family of transistors: alloy, alloy diffused, diffused mesa, diffused field effect, epitaxial, planar, grown junction, surface barrier, and many other variations. We will therefore concern ourselves with this family, and neglect the point-contact type.

7 1. THE PRINCIPLE OF THE JUNCTION TRANSISTOR

The two types of junction transistor are *p–n–p*, in which con duction predominates in *p*-type material, and *n–p–n*. The principle of operation is basically the same, so it will suffice to consider only the *p–n–p*. To start with we will consider the *p–n–p* transistor as a thin layer of *n*-type material between two blocks of *p*-type material. This gives us two junctions; the first piece of *p*-type material is called the *emitter*, the *n*-type is the *base*, and the other *p*-type is called the *collector*. Connections are brazed or soldered to the three regions as shown in Figure 7.1.

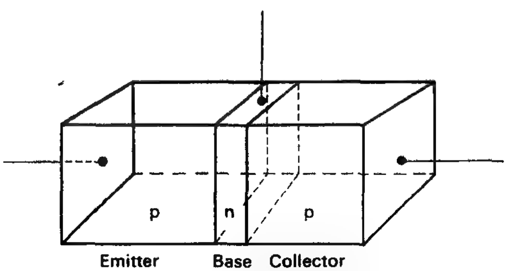


FIG. 7.1. The basic *p–n–p* transistor

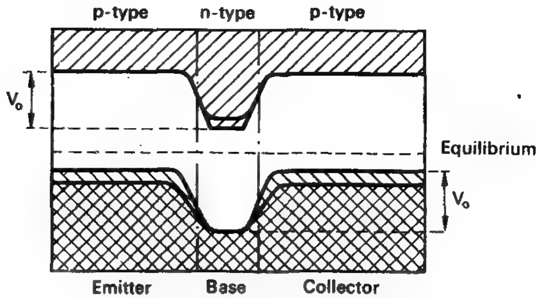


FIG. 7.2. The *p–n–p* transistor in equilibrium

Figure 5.8(a) showed the energy band model for a single *p–n* junction in equilibrium. Likewise Figure 7.2 represents the two "back-to-back" *p–n* junctions of the transistor, also in equilibrium. The essential point about the operation of the transistor is that the emitter-base junction is forward biased whilst the collector-base junction is reverse biased (see Figure 7.3). The forward-biased junction injects majority carriers (in this case holes) into the base region. They diffuse through this, and most reach the reverse-biased junction and are attracted into the collector (hence its name). Some of the holes injected into the base recombine with electrons and are lost as far as the process of conduction between emitter and collector is concerned.

Each hole reaching the collector from the base requires an electron from the external supply to neutralize it, and this flow of electrons into the collector from outside is called the collector current (I_c). The collector junction (reverse biased) injects electrons into the base where they are majority carriers (because the base is of *n*-type material). The necessary balance is maintained from another external supply connected to the base, giving rise to a base current I_B. Application of Kirchhoff's Law shows that the emitter current I_E must be given by

$$I_E = I_C + I_B \tag{7.1}$$

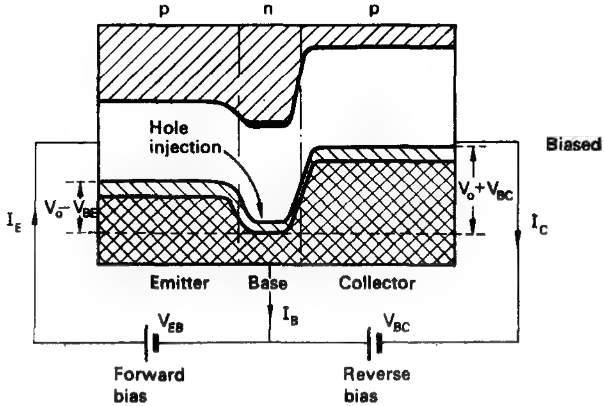


FIG. 7.3. The *p–n–p* transistor with its emitter-base junction forward biased; and its base-collector junction reverse biased

Since I_C is less than I_E, at first sight it may seem odd that the device is any use at all, until one remembers that I_E is associated with a low resistance (forward bias), whilst I_C passes through a high resistance (reverse bias). So a small power change in the emitter-base circuit can control a large power change in the base-collector circuit. For example, a typical transistor may have an emitter-base resistance of 500 Ω, and a base-collector resistance of 500,000 Ω; so if the current through each junction is virtually the same, the power gain is 1000 : 1 and the voltage gain similar.

Clearly I_B must be kept as small as possible, so that I_C is nearly equal to I_E. This is achieved by careful design of the impurity concentrations in the three regions, and by careful production control.

To help understand the transistor's action we give a specific example of a typical *p–n–p* germanium structure, for which it is decided the conductivities of the emitter, base, and collector regions will be 10^4, 100, and 10 mho m^{-1} respectively.

We saw in eqn. (1.13) that

$$\sigma = ne\mu$$

where σ = electrical conductivity,
n = no. of charge carriers per unit volume, or density,
e = charge on electron or hole ($1{\cdot}6\times10^{-19}$ coulomb),
and μ = mobility of charged carrier.

We also have eqn. (3.1):

$$np = k$$

where n = density of free electrons (as above), p = density of holes, and k is a temperature-dependent constant.

If we use i_c to stand for intrinsic carrier density (the density of holes and electrons intrinsically present, taken together), the constant k in (3.1) (above) is equal to i_c^2, so we have

$$np = i_c^2 \tag{7.2}$$

At room temperatures the respective mobilities of electrons and holes are 0·36 and 0·17 $m^2\ V^{-1}\ sec^{-1}$. So, if the emitter is to have a conductivity of 10^4 mho m^{-1} and an intrinsic carrier density of $2{\cdot}5\times10^{19}\ m^{-3}$, we obtain from (1.13):

$$10^4 = n(1{\cdot}6\times10^{-19}\times0{\cdot}36)+p(1{\cdot}6\times10^{-19}\times0{\cdot}17) \tag{7.3}$$

and

$$np = (2{\cdot}5\times10^{19})^2 \tag{7.4}$$

Solving (7.3) and (7.4) gives

$$p = 3{\cdot}68\times10^{23}\ m^{-3}$$

and

$$n = 1{\cdot}70\times10^{15}\ m^{-3}$$

The emitter is a *strongly* doped *p*-type material.

A similar calculation for the base gives:

$$p = 3{\cdot}57\times10^{17}\ m^{-3}$$

and

$$n = 1{\cdot}75\times10^{21}\ m^{-3}$$

Thus the base is a *lightly* doped n-type material.
For the collector:

$$p = 3{\cdot}68 \times 10^{20}\ \mathrm{m}^{-3}$$

and
$$n = 1{\cdot}70 \times 10^{18}\ \mathrm{m}^{-3}$$

This is a *lightly* doped p-type material.

These results are summarized in Table 7.1.

TABLE 7.1

	Emitter	Base	Collector
Majority carrier concentration (per m³)	$3{\cdot}68 \times 10^{23}$ holes	$1{\cdot}75 \times 10^{21}$ electrons	$3{\cdot}68 \times 10^{20}$ holes
Minority carrier concentration (per m³)	$1{\cdot}70 \times 10^{15}$ electrons	$3{\cdot}57 \times 10^{17}$ holes	$1{\cdot}70 \times 10^{18}$ electrons
Electrical conductivity (in mho m⁻¹)	10^4	10^2	10^1

If in Figure 7.3 the forward-bias battery in the emitter-base circuit is disconnected, the emitter carries no current and the emitter-base junction is left with the "built-in" junction p.d. V_o (see Section 5.4 and Fig. 5.8(a)), and the hole current components are equal and opposite. The electron current components (I_D and I_G in Fig. 5.4) are also equal and opposite, but are negligible

anyway because the hole concentration in the emitter (3.68×10^{23} m^{-3}) is far greater than the electron concentration in the base (1.75×10^{21} m^{-3}). The only action then is at the base-collector junction, which is reverse biased. The potential barrier height is $V_o + V_{BC}$, the depletion layer extending almost right through the base ($\sigma = 100$ mho m^{-1}) but hardly into the collector ($\sigma = 10$ mho m^{-1}). As we saw in Section 5.4, the collector draws the saturation current I_G which is due to thermally generated holes in the base diffusing to the collector. The potential barrier is too high for any electrons to pass from base to collector, but electrons can flow from collector to base. This is, however, small, because there are few electrons ($1{\cdot}70 \times 10^{18}$ m^{-3}) in the collector. This saturation current I_G is called the "collector cut-off current" in transistor technology, and given the symbol I_{co}. (Such a "leakage" current after *cut-off* has taken place is not to be found in the corresponding thermionic triode situation.)

When the forward bias V_{EB} (see Fig. 7.3) is reconnected, the potential barrier at the emitter-base junction is reduced in height by V_{EB} from V_o to $V_o - V_{EB}$, allowing holes from the emitter to "spill over" into the base. These holes rapidly diffuse more or less uniformly throughout the base region. Most eventually reach and pass over the "easy" base-collector junction; some are lost by recombination with intrinsic free electrons in the base. As the concentration of electrons in the base is only $1{\cdot}75 \times 10^{21}$ m^{-3}, these recombinations are few; but they do require a "base current" of electrons from the external supply into the base. (Again, in normal operation the grid of a thermionic triode passes *no* current.)

As the injection of holes from emitter to base is controlled by V_{EB}, a small change in this variable will produce a relatively large change in I_C. On this fact depends the transistor's ability to "amplify".

7.2. THE CURRENT GAIN, α

Usually the "gain" of an amplifier should be as high as possible, i.e. the collector current I_C should be as high a fraction of the emitter current I_E as possible. This leads to two important considerations: (1) the emitter-base current should consist mainly of holes injected to the base from the emitter, as the electrons passing from base to emitter contribute to I_E but not I_C, reducing the desired ratio. This loss is termed the *emitter efficiency*, γ, ("gamma")

$$\text{where } \gamma = \frac{\text{Hole current from emitter to base}}{\text{Total (holes+electrons) current between emitter and base}}$$

γ is only slightly less than unity because the emitter hole concentration is much greater than the base electron concentration (see the figures above).

It can be shown that

$$\frac{I_{Eh}}{I_{Ee}} = \frac{\sigma_E}{\sigma_B} \times \frac{Le}{w} \tag{7.5}$$

where I_{Eh} = emitter-to-base *hole* current, I_{Ee} = base-to-emitter *electron* current, Le = diffusion length (see below) of electrons in the emitter region, and w = width of base region.

(The diffusion length Le is the length over which the effect of the excess carriers is "noticeable", and in this example is a measure of the depth into the emitter material over which the density of injected electrons has been reduced by a factor $1/e$.)

The ratio $\dfrac{\sigma_E}{\sigma_B} = \dfrac{10^4}{100} = 100$, and $\dfrac{Le}{w} \triangleq \dfrac{10^{-3}\text{ m}}{10^{-5}\text{ m}} = 100.$

So, in eqn. (7.5), $\dfrac{I_{Eh}}{I_{Ee}} \triangleq 100 \times 100 = 10^4$

and $$\gamma = \frac{I_{Eh}}{I_{Eh}+I_{Ee}} = \frac{1}{1+10^{-4}} \simeq 1.$$

The second consideration for high "gain" is that as many holes as possible, passing through the base, should reach the collector without recombining with electrons. The loss due to those that *do* recombine is expressed by the *transport factor*, β ("beta"), where

$$\beta = \frac{\text{hole current into collector}}{\text{hole current out of emitter}}$$

It can be shown that the fraction of holes lost in the base by recombination is $\frac{1}{2}(w/L_h)^2$, where L_h is the diffusion length for holes in the base region (about 10^{-3} m), so

$$\beta \simeq 1 - \frac{1}{2}\left(\frac{10^{-5}}{10^{-3}}\right)^2 \simeq 1$$

Notice that in the efforts to make γ and β as near to unity as possible, we have produced another condition: that the width of the base should be very thin. This in turn has another advantage: holes from the emitter diffuse fairly slowly through the base to the collector, because there is no electric field across it. The time it takes holes to cross the base limits the high-frequency use of the transistor (cf. transit time in a thermionic triode).

Conflicting considerations apply to the junction *areas*. A small area makes for a small capacitance and better high-frequency performance, whereas it limits the power-handling capability and adversely affects the transport factor.

Yet another factor affects the current gain α. It is possible for holes sweeping through the collector at high collector base voltages to give rise to secondary holes and electrons by a process called *collector multiplication*. (In thermionics, cf. secondary emission.) Carriers so produced add to the collector current and

increase the ratio I_C/I_E. We define *collector efficiency* δ ("delta") as

$$\delta = \frac{\text{total collector current}}{\text{incident hole current}}$$

The cumulative effect of β, γ, and δ is summarized in the one parameter α ("alpha"), the *current amplification factor*, defined by the relation

$$\alpha = \left(\frac{\partial I_c}{\partial I_E}\right)_{V_{BC}}$$

or

$$\frac{\text{change in collector current}}{\text{change in emitter current}}$$

at constant base-collector voltage. In practice α is not strictly constant over the full current range of any transistor; departures from the nominal value are quite marked at low and high currents. α may be greater or less than unity, and is (in the mathematical sense) a complex number. But in the absence of collector multiplication α is always less than unity, and typically lies between 0·900 and 0·995—the higher the "better".

Simply then we can establish a relation between I_C (and its associated "leakage current" I_{co}) and I_E. We have

$$I_C - I_{co} = aI_E$$

and

$$I_C = \alpha I_E + I_{co} \tag{7.6}$$

A graph of I_C against I_E is shown in Figure 7.4.

Eliminating I_C from eqns. (7.1) and (7.6), we have

$$I_B = (1-\alpha)I_E - I_{co} \tag{7.7}$$

The explanation of the *n–p–n* transistor is the same as the foregoing, except that supply polarities are reversed, and holes and electrons interchanged.

In Chapter 8 we go on to discuss the construction and characteristics of transistors.

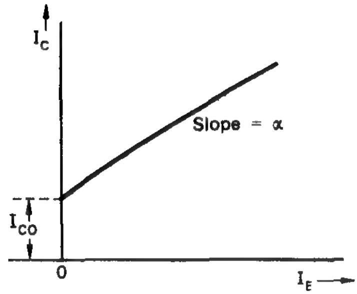


FIG. 7.4. Graph of collector current/emitter current for a common base transistor configuration

QUESTIONS

1. Draw careful diagrams to correspond with those of Figures 7.1 to 7.4, for an *n–p–n* transistor.

2. Define and explain the terms emitter efficiency, diffusion length, transport factor, collector multiplication. Explain what is meant by the statement that the current amplification factor is a complex number.

3. Discuss the considerations which limit (a) the magnitude, and (b) the frequency, of the signal which can usefully be amplified by a transistor.

4. Sketch energy band diagrams for (i) an unbiased, and (ii) a normally operating, *p–n–p* transistor; show also the hole concentration across the device during normal quiescent operation and relate this diagram to the collector current.

Explain the terms emitter efficiency and transport efficiency, and outline the physical parameters in the device which control their values. Discuss qualitatively the factors that govern the choice of the impurity concentrations for the base and collector regions.

(U.L. B.Sc. (Eng.) Part 2: Electrical, 1967)

5. Explain the operation of an *n–p–n* transistor, indicating in particular the meaning of the terms valency and junction. Why is a transistor temperature-sensitive?

What will happen to an *n–p–n* transistor connected as an amplifier if its base is shorted to its emitter?

(C. and G.: Advanced Telecomms and Electronics,
Digital Computers, May 1967)

CHAPTER 8

TRANSISTOR PRODUCTION AND CHARACTERISTICS

WE DISCUSSED in some detail the methods of *p–n* junction formation in Section 6.1, and here we will briefly extend the principles to transistor production. Then we will discuss some of the more important electrical characteristics of the transistor.

8.1. TRANSISTOR PRODUCTION

We saw in Chapter 7 that a transistor consists of three different regions in a piece of semiconductor material, separated by two *p–n* junctions. The regions may have different electrical conductivities, areas, and thicknesses, and the piece of material may be basically *p*-type (for a *p–n–p* transistor) or *n*-type. Connections are made to the three regions and the whole encapsulated. Table 8.1 lists data of some typical transistors.

8.1.1. The germanium alloy transistor

The structure is shown in Figure 8.1. The process of manufacture begins with two pellets of indium (say), which are alloyed to opposite faces of a thin *n*-type germanium wafer to form emitter and collector. A nickel base tab is then fused to the wafer to support the structure and make the base connection. The alloy-

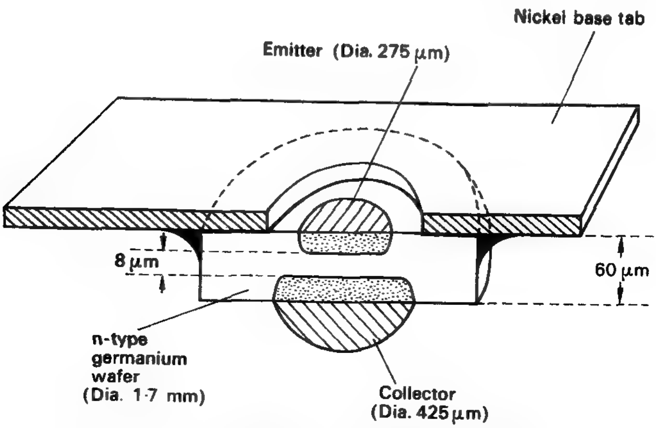


FIG. 8.1. A germanium alloy transistor element.
(Courtesy of R. G. Hibberd)

ing process (described in Section 6.1.1) is very carefully controlled so that the resulting base width is about 10 μm (or about half a thousandth of an inch). Note that the collector area is larger than the emitter area; this is to keep the transport factor (and consequently the current gain) as near unity as possible.

The nickel base tab, with transistor attached, is then mounted to a "header" which is a disc with three insulated through leads (see Fig. 8.2). Two small wires are welded one each to a through lead, and their other ends are fused to the emitter and collector regions. The nickel base tag is welded to the third through lead.

The production process is now effectively complete, except for a few precautionary measures. To avoid surface contamination, which might shunt the junctions or diffuse into the material and form energy levels between the conduction and valence bands (see Section 3.4), the unit is treated electrolytically and then thoroughly washed in pure water. It is dried by prolonged

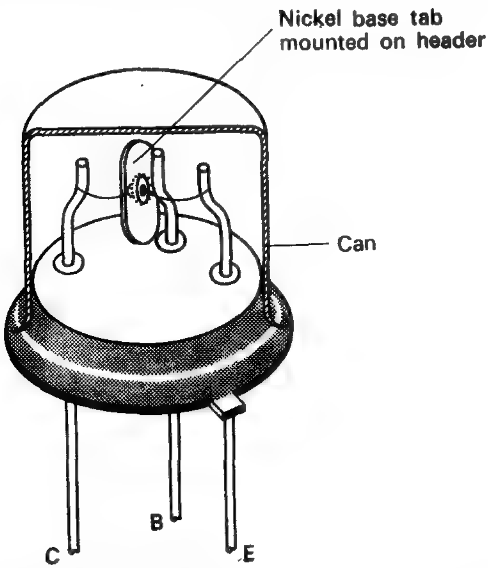


FIG. 8.2. The germanium alloy transistor element mounted on its "header" and installed in its can. (Courtesy of R. G. Hibberd)

baking at about 110°C, and then encapsulated by welding a metal can over the header. Silicone grease or varnish are sometimes used inside the can to improve surface protection and thermal conduction from the collector region to the outside. The case also excludes light.

A typical low power germanium alloy transistor is specified in line (b) of Table 8.1.

TABLE 8.1

Type	Material	(Dimensions 25 μm ≃ 1/1000″) Emitter (μm)	Collector (μm)	Base width (μm)	Cut-off frequency (MHz)	V_C (V)	I_C (A)	P_C (W)	Junct. temp. (°C)	I_{CO} at 20°C (μA)	C_{BC} at 6 V (pF)
(a) Alloy	Silicon	825 dia.	1100 dia.	13	1	40	0·1	0·3	200	0·01	40
(b) Alloy	Germanium	275 dia.	425 dia.	8	10	25	0·2	0·2	85	1	20
(c) Alloy Power	Germanium	3000 dia.	4500 dia.	18	0·4	60	7	≃150	100	200	200
(d) Alloy diffused	Germanium	125 dia.	150×300	3	150	40	0·03	0·1	85	1	12
(e) Diffused mesa	Germanium	75×25	100×100	1	800	30	0·05	0·1	85	1	2
(f) Diffused mesa (small signal)	Silicon	150×50	200×200	$1\frac{1}{2}$	250	60	0·1	0·3	200	0·01	8

(g) Epitaxial planar (small signal)	Silicon	125 dia.	250 dia.	$1\frac{1}{2}$	300	80	0·1	0·3	200	0·001	8
(h) Epitaxial planar (medium power)	Silicon	375 dia.	800 dia.	2	50	80	0·5	0·8	200	0·01	20
(i) Epitaxial planar (high power)	Silicon	"E" shape 3800 × 2500	5000 × 5000	2	10	80	7	$\simeq 50$	200	0·1	500

The advantages of the alloy junction transistor include low resistance between the junctions and external leads (giving a low voltage drop across the device when saturated), high reverse emitter breakdown voltage, cheap and easily automated production. Unfortunately the production of very thin base regions is difficult, and such transistors have an upper frequency limit of about 10 MHz. There is an upper limit also to the collector voltage which, if exceeded, will cause a Zener (q.v.) effect; this limit arises because the electrical conductivity of the recrystallized layer in the collector is higher than that in the base material, so the collector depletion layer moves mainly into the base region.

8.1.2. The silicon alloy transistor

The structure is shown in Figure 8.3. Although generally similar to its germanium counterpart, a different construction is necessary. Aluminium is alloyed into *n*-type silicon but the

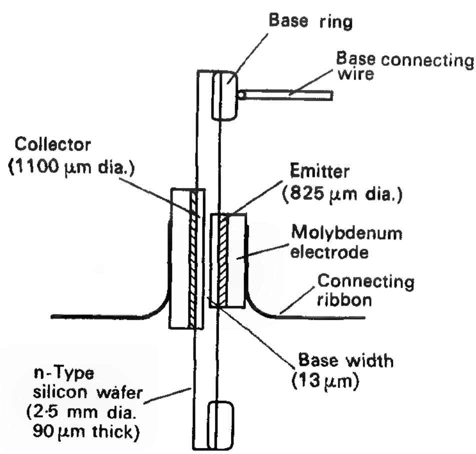


FIG. 8.3. A silicon alloy transistor element. (Courtesy of R. G. Hibberd)

alloy formed is quite brittle and has a different coefficient of thermal expansion from silicon. Thus a strain can be set up in the silicon after the alloying process is over and whilst solidification is taking place. (This is not important in the germanium case because the indium used is quite soft.)

Comparison of lines (a) and (b) in Table 8.1 show the advantages associated with silicon alloy transistors: low collector leakage current, higher operating temperature limit and consequently ability to handle more power, and inexpensive to make. However, the cut-off frequency (1 MHz) is lower than for germanium, because the electron and hole mobilities in silicon are less than in germanium at room temperatures (0·17 and 0·035 $m^2 V^{-1} sec^{-1}$ in silicon, 0·38 and 0·18 in germanium).

8.1.3. The germanium alloy power transistor

Typical data appears in line (c) of Table 8.1. The power handling capacity is about 10^3 times that of normal alloy transistors, and although the basic alloying process is the same, the construction (Fig. 8.4) is different to allow for far greater power dissipa-

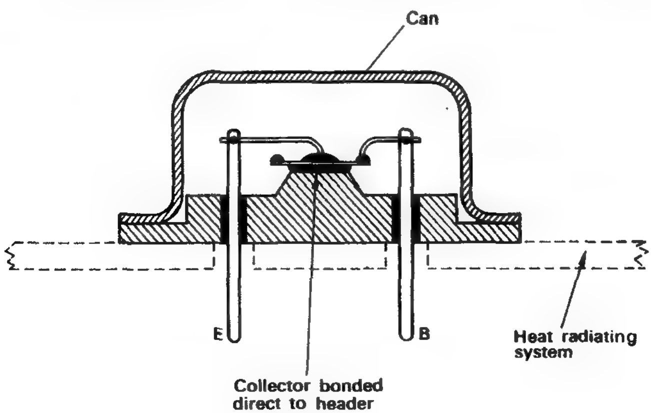


FIG. 8.4. Cross-section of a germanium alloy power transistor

tion. Heat is dissipated by bonding the collector directly to the header and, if still necessary, externally fixing the collector to some kind of "heat sink".

8.1.4. The germanium diffused mesa transistor

As we saw in Section 6.1.2, the diffusion process gives finer control over base width and junction area than does the alloy process. This is most important because the transit time of the minority carriers through the base regions is one limit to the high-frequency response, and the mobility of carriers is higher in germanium than in silicon (Sect. 8.1.2), so germanium is pre-

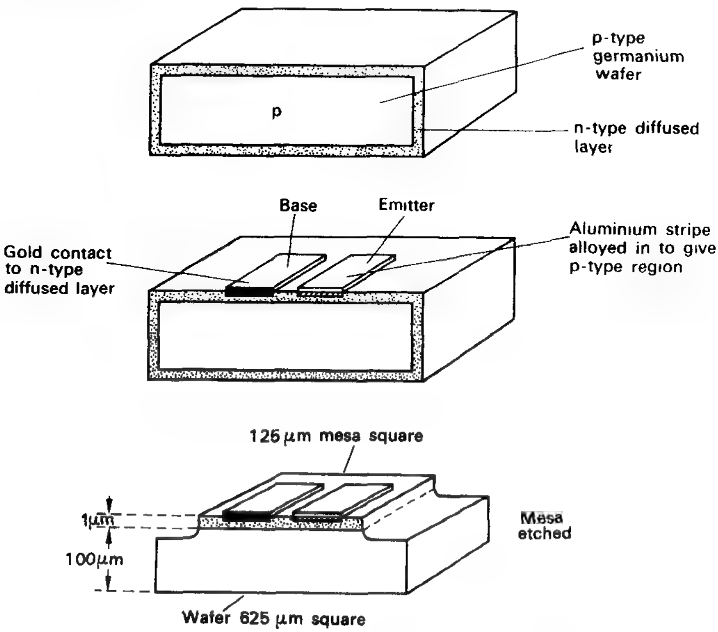


FIG. 8.5. Construction of a germanium diffused base transistor. (Courtesy of R. G. Hibberd)

ferred for frequencies of about 1 GHz. Some typical data appears in line (e) of Table 8.1. The stages of construction are shown in Figure 8.5, and the base width will typically be about 1 μm. The region of the diffused collector is defined by masking over the emitter and base regions and etching the wafer to leave a "mesa" shape (the name comes from American desert rock outcrops of similar shape) which contains the two junctions.

A very important production feature is that the process so far described can be carried out on many transistors simultaneously: indeed a germanium slice of about 2 cm diameter can yield in one step 1000 such transistors. The slice is then cut into the individual transistors, and gold wires bonded to the emitters and bases by thermo-compression techniques involving pressures of some 2×10^7 N m^{-2} and temperature of 300°C.

The alloy diffused transistor (line (d) of Table 8.1) uses a combination of alloying and diffusing techniques. Base thickness is about 3 μm and cut-off frequency about 150 MHz.

8.1.5. The silicon diffused mesa transistor

A recent type is shown in Figure 8.6, and data appears in Table 8.1, line (f). The base and emitter layers (in an *n–p–n* type) are formed by separate donor and acceptor diffusions. When the

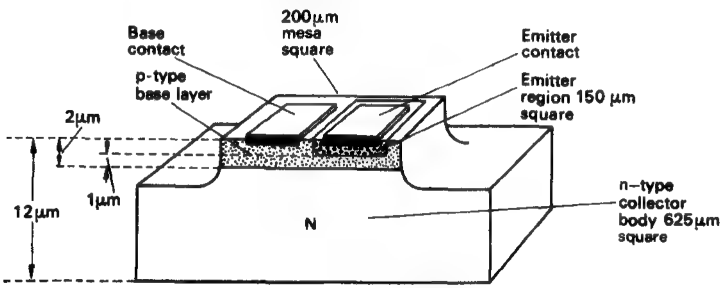


FIG. 8.6. A silicon diffused mesa transistor.
(Courtesy of R. G. Hibberd)

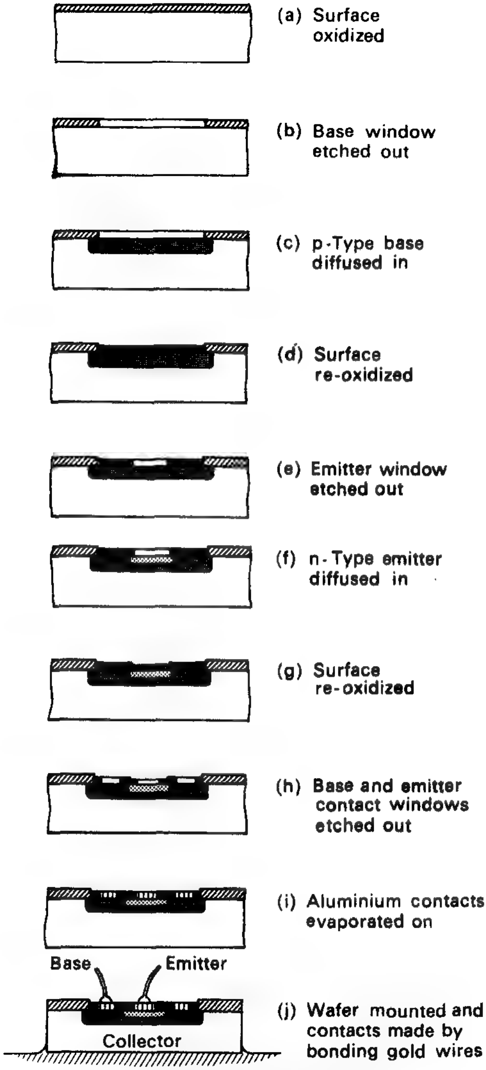


FIG. 8.7. The silicon diffused planar process. (Courtesy of R. G. Hibberd)

p-type base layer has been diffused, the surface of the silicon slice is oxidized, and a photographic process used to etch windows in the oxide. (The process is called "photo-resist", because areas which have received a suitable optical exposure will thereafter resist the etching compound.) The emitter is then formed by diffusing phosphorus through the window, diffusion *not* occurring elsewhere because of the oxide surface. The collector junction is then defined by mesa etching. The whole unit is thoroughly washed and dried before mounting, great care being taken to keep the base–collector region surfaces free from contamination, as this would increase the leakage current I_{co}. It was research work into methods of protecting the surface that led to the "planar" process next to be described.

8.1.6. The silicon epitaxial planar transistor

This most important process, many steps of which are similar to those used in the diffused mesa process, is illustrated in Figure 8.7, and a large number of these transistor elements—perhaps 300 from one 2 cm diameter piece of silicon—are produced simultaneously.

Figure 8.8 shows a typical element capable of working in the 50–100 MHz range with a collector current of about 200 mA; it would be cut from the 2 cm slice and mounted on a header in the usual way (Fig. 8.9).

The important step in the process is the protection of the base-collector junction surface by the oxide layer, resulting in very low and stable collector leakage currents (Table 8.1, lines (g), (h), (i)). Also, because the process is basically one of diffusion, very thin base widths giving cut-off frequencies of several GHz are possible.

From the production point of view the same basic process and equipment can produce a variety of different size transistors,

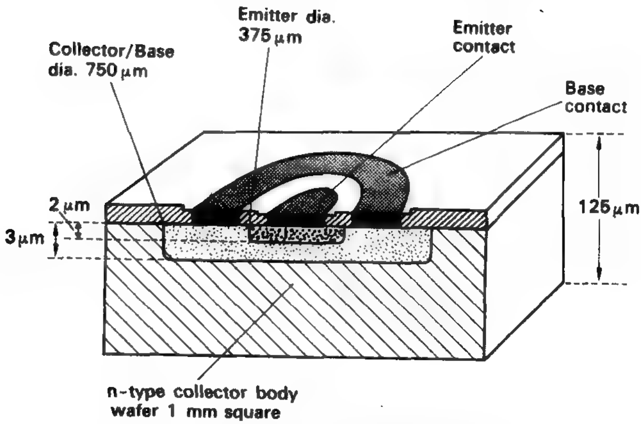


FIG. 8.8. A silicon planar transistor wafer.
(Courtesy of R. G. Hibberd)

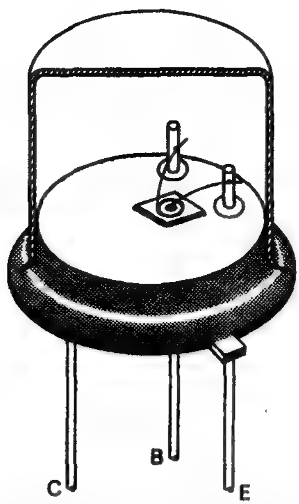


FIG. 8.9. The silicon planar wafer on its header and in its can.
(Courtesy of R. G. Hibberd)

by substituting different photographic masks in the photo-resist process. In addition, although the process consists of several steps, up to 300 are produced simultaneously so the cost need not be prohibitive.

The main fault with the process is the relatively high value of resistance between the collector junction and the lead, and of the contact at the bottom surface of the wafer. As the collector

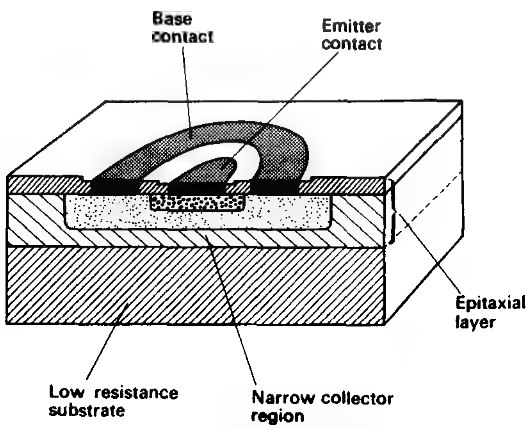


FIG. 8.10. A silicon epitaxial planar transistor.
(Courtesy of R. G. Hibberd)

current flows through this resistance a p.d. is set up which results in a high value of saturation voltage. The resistance of this region can be reduced by using a composite slice, consisting of a low-resistance substrate on which is the thin epitaxial layer of collector material in which the transistor is formed (Fig. 8.10). Such a transistor is called epitaxial *planar*, and is a versatile device with good all-round characteristics.

8.2. TRANSISTOR CHARACTERISTICS

We saw in Chapter 7 that the transistor is capable of amplification. Before amplifiers can be properly designed (given output power for given input signal, with frequency range and distortion specified), the electrical characteristics of the transistors, as used in whatever particular method of connection, need to be known.

8.2.1. Common base characteristics

If the transistor is connected to the two voltages as shown in Figure 7.3, the base is common to both voltages and this method of connection is called "common base". Figure 8.11 shows a common base amplifier circuit with the signal applied in series with

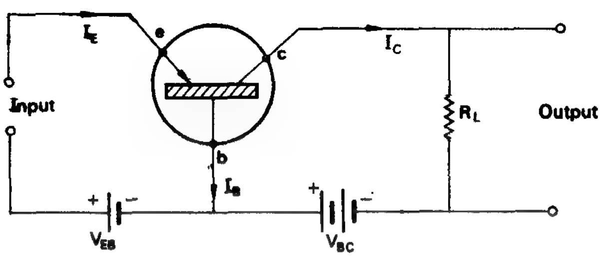


FIG. 8.11. *p–n–p* transistor connected as a common base amplifier

the emitter-base battery V_{EB} and with a load resistor R_L connected in series with the base-collector battery V_{BC} (which is, of course, reverse biased). Note the circuit symbol for a *p–n–p* transistor. The symbol for an *n–p–n* transistor is similar but the arrow distinguishing the emitter is reversed; this arrow

always points in the "easy" direction of conventional (+ to −) current flow (see Fig. 8.12).

The "input characteristic" of the common base circuit is a graph showing how the emitter current I_E increases as the base–emitter voltage V_{EB} is increased (see Expt. 8.1). It is the same as the characteristic of a forward-biased *p–n* junction and is shown in Figure 8.13 (compare with Fig. 5.7). The scale shown is typical of a small germanium transistor, and assumes a con-

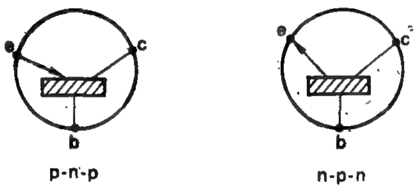


FIG. 8.12. Transistor circuit symbols

stant value of collector–base voltage V_{CB}. The effect of varying V_{CB} is slight (compare the linear relationship between I_C and I_E shown in Fig. 7.4) but for a given value of V_{EB} the emitter current I_E rises slightly as the reverse bias voltage V_{CB} across the collector–base junction rises. This phenomenon, the effect of collector voltage on the hole density gradient at the emitter junction, was explained in 1952 by Early as being due to base widening.

The "output (or collector) characteristic" in the common base configuration is a graph showing the variation of collector current I_C with the collector voltage V_{CB}: for zero emitter current it is just that of the reverse-biased *p–n* junction, and is shown in Figure 8.14. (Compare this also with Fig. 5.7.) This small current is the leakage current I_{CO}. If now an emitter current of 1 mA flows (point *A* in Fig. 8.13), the output characteristic will be the curve *A* of Figure 8.14. This shape arises because nearly all the emitter current flows into the collector but the current

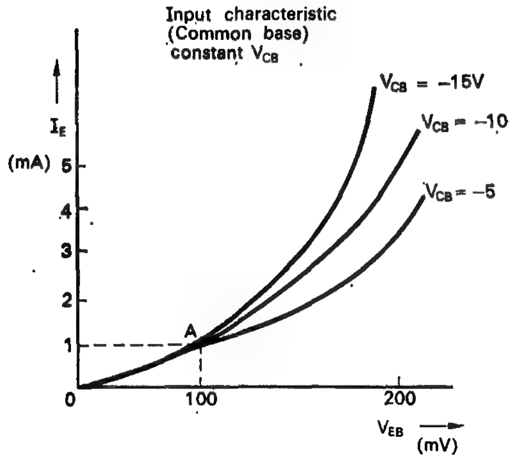


FIG. 8.13. Transistor input characteristic (in common base)

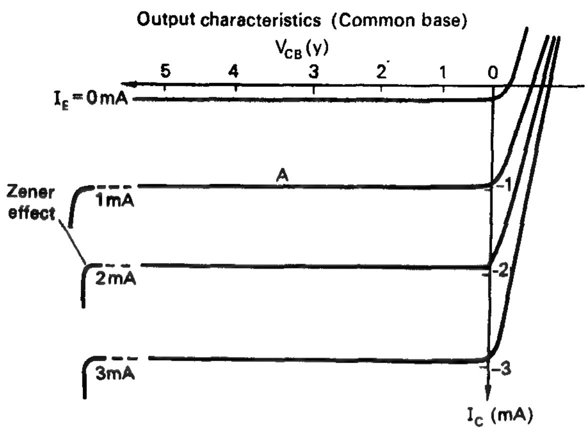


FIG. 8.14. Transistor output characteristic (in common base)

collected is ***not*** dependent on the collector voltage. Similarly we have the curves for $I_E = 2, 3, 4$ mA, etc. Notice that for a given value of I_E, I_C remains sensibly constant right down to zero collector voltage. This is explained by the small base current flowing out of the resistance of the base region creating a small p.d. which acts as a small reverse bias across the base–collector

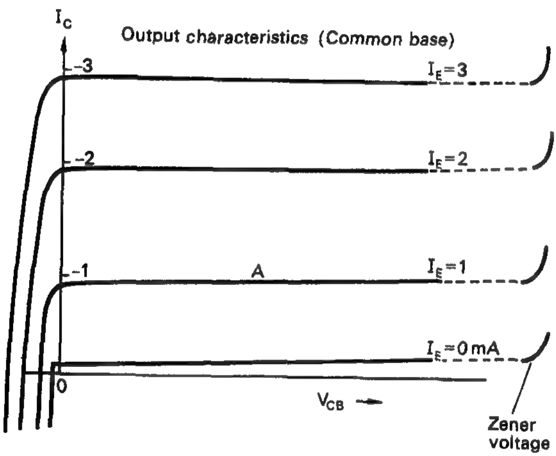


Fig. 8.15. Figure 8.14 rotated through 180°

junction. (In order to reduce the collector current to zero a small *forward* bias on the base–collector junction would be necessary.) We have drawn Figure 8.14 in the "third quadrant" for clarity; it is usual to rotate this 180° as shown in Figure 8.15. The student should now notice the similarity to the anode current/anode voltage characteristic of a vacuum *pentode* valve (see Vol. 1, Fig. 7.9). The output resistance can be seen to be very high, a large change of V_{CB} produces only a small change of I_C. Note in Figures 8.14 and 8.15 that if V_{CB} exceeds the breakdown voltage a Zener (q.v.) effect occurs.

8.2.2. Common emitter characteristics

In the common emitter configuration the input is applied between the base and emitter, and the output is taken from the collector and emitter circuit (see Expt. 8.2). Note in Figure 8.16 that the emitter is positive with respect to the base, i.e. the emitter–base junction is forward biased; and that the collector is negative with respect to the base (reverse biased). Note also,

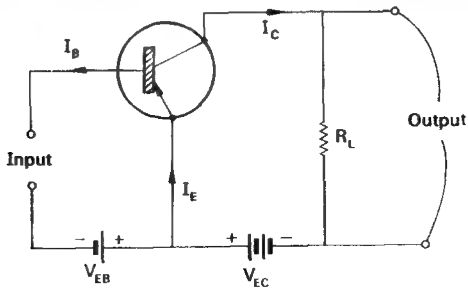


FIG. 8.16. *p–n–p* transistor connected as a common emitter amplifier

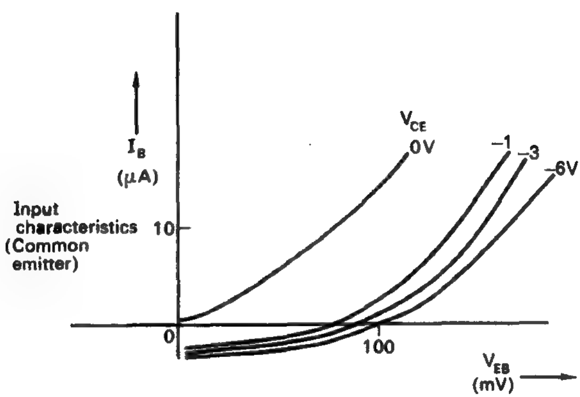


FIG. 8.17. Transistor input characteristic (in common emitter)

however, that the collector voltage V_{EC} is now applied relative to the emitter instead of the base.

The "input characteristic" of the common emitter circuit is the graph showing how base current I_B increases as the voltage across the emitter–base junction increases in the forward direction. Figure 8.17 shows a set of such characteristics for various

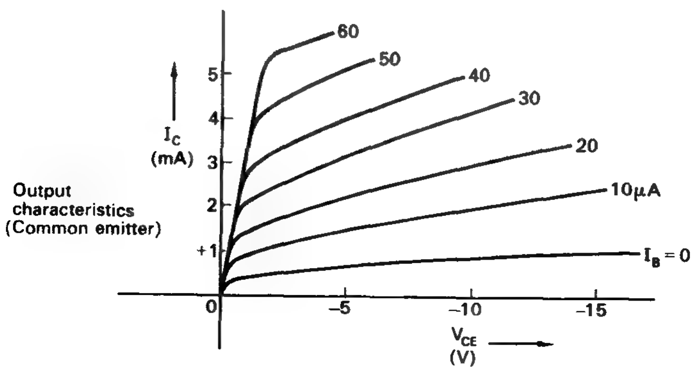


FIG. 8.18. Transistor output characteristic (in common emitter)

fixed values of V_{EC}. The family of curves is displaced from the zero current axis by an amount equal to the leakage current I_{co}, which is a component of the base current. Since the current gain increases with collector voltage, the value of I_B for a particular value of V_{EB} decreases for increase in V_{CE}—the Early effect again. By comparing Figures 8.17 and 8.13 it can be seen that the input resistance (the reciprocal of the slope of the graph) in common emitter configuration is higher than in common base.

The output (or collector) characteristic is a graph showing the variation of collector current I_C with collector–emitter voltage V_{CE} for various values of base current I_B (see Fig. 8.18).

Points to notice on this very important characteristic include:

(a) At constant V_{CE} a small change I_B (for example 10 μA) produces a very much larger change in I_C (perhaps 1000 μA), i.e. a gain of 100 : 1.

(b) The curve $I_B = 0$ is displaced considerably from the V_{CE} axis.

(c) The "knee" of each I_B curve is displaced to the right as I_B increases.

(d) Comparing Figures 8.18 and 8.15, it is clear that the output resistance (the reciprocal of the slope of the graph) is considerably *less* for the common emitter circuit than for the common base.

In Chapter 9 we shall develop an "equivalent circuit" for the transistor and use it to explain many of the apparently mysterious phenomena of transistor characteristics. We will go on to consider transistors as amplifiers of electric signals. (The student may note that a third form of connection, common collector, exists; but we will not be concerned with it in this volume.)

EXPERIMENT 8.1: "Common base" characteristics.

Apparatus needed:

D.C. supplies, 1 V, 200 mA. (These must be separate.)
2 potentiometers, 15 Ω 1 W.
2 voltmeters, 1 V f.s.d. ("Unilab" 1 mA 100 Ω basic meters are ideal).
2 milliammeters, 2–0–10 mA (Nuffield pattern).
1 transistor type OC81.

Procedure. Set up the circuit of Figure 8.19. Attach paper labels to each meter to show at a glance which reads which: V_{CB}, I_C, I_E, and V_{EB}. (This is sure to save a lot of time.)

The "input characteristic" is a graph of I_E against V_{EB}, or a set of such graphs taken at different fixed values of V_{CB}. When V_{EB} is changed, V_{CB} also changes and may have to be reset before the value of I_E can be taken.

The "output characteristic" is a set of graphs of I_C against V_{CB}, I_E being kept constant. The interaction is stronger this way round, and each time

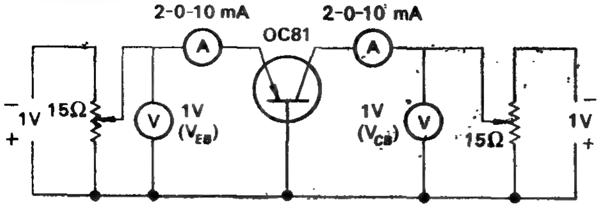


FIG. 8.19.

a new value of V_{CB} is set it will be necessary to readjust V_{EB} to re-establish the desired value of I_E. Luckily, this has very little effect on the new value of V_{CB}, which will probably not need further adjustment.

Precaution. If none of the meters are allowed to go off scale, the transistor will be running well within its capabilities.

Question. The low-resistance potentiometers take far more current than the transistor. Why do you suppose high-resistance potentiometers cannot be used? (It is safe to try this.)

EXPERIMENT 8.2: "Common emitter" characteristics.

Apparatus needed:

Transistor type OC81.
Milliammeter, f.s.d. 50 mA.
Milliammeter, f.s.d. 1 mA.
Voltmeter, f.s.d. 1 V.
Voltmeter, f.s.d. 5 V.
D.C. supplies, 1 V and 5 V.

Procedure. Set up the circuit of Figure 8.20 and label each meter as before.

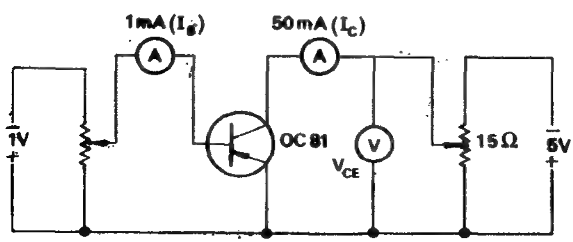


FIG. 8.20.

To graph the input characteristics, obtain values of I_B and V_{EB} by varying V_{EB}, whilst keeping V_{CE} constant. Keep an eye on the I_C meter, and do not allow it to go off the scale.

The output characteristics are graphs of I_C against V_{CE} for various fixed values of I_B. (A reading of V_{BE} is not needed, and this meter may be removed.) After setting a new value of V_{CE}, it will usually be necessary to reset I_B to the chosen value.

Precaution. As before, if the meters do not go off the scales, the transistor will not be exceeding its ratings.

QUESTIONS

1. Why does a silicon alloy transistor have an inherently lower maximum usable frequency than a germanium alloy type? What are the advantages of the silicon type?

2. Sketch the circuits you would use for measuring (a) the common base, and (b) the common emitter, characteristics of an *n–p–n* transistor. Quote suitable values for the components, supplies, and measuring instruments you would use.

Sketch a graph of a typical common base input characteristic, putting in suitable figures, and determine from the graph the "input resistance" of the example you have sketched.

3. What is meant by the "Early effect", and how do you account for it? In what circumstances is it a disadvantage?

4. Draw and explain a circuit diagram showing the basic form of a common-emitter transistor amplifier. Compare the action of a transistor amplifier with that of a vacuum triode amplifier.

(U.L.C.I., Electrical Technology, June 1966)

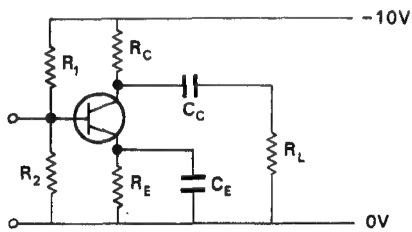


FIG. 8.21.

5. Explain the function of the various components of the common-emitter transistor amplifier shown in Figure 8.21.

(U.L. B.Sc. (Eng.) II, 1967; part question)

6. Explain in detail, with the aid of diagrams, the action of a germanium *n–p–n* transistor. Describe the main processes in the manufacture of germanium junction transistors.

(U.L. B.Sc. (Eng.) II, 1967)

7. Discuss the factors which affect the inherent frequency response of different types of transistor.

(Grad. Inst. P, 1966, II)

CHAPTER 9

THE TRANSISTOR AS A CIRCUIT ELEMENT

9.1. A LOW-FREQUENCY EQUIVALENT CIRCUIT FOR COMMON BASE

We saw in Section 7.2 that

$$I_c = \alpha I_e + I_{co} \tag{7.6}$$

which, if we neglect the leakage current I_{co} and the variation of α, shows that a fraction α of the emitter current I_e reaches the collector circuit as the collector current I_c.

In Section 5.5 we showed that the resistance of a forward-biased *p–n* junction is $25/I$ ohms approximately, where I is in mA. In our transistor this represents the resistance between emitter and base and we call it r_e.

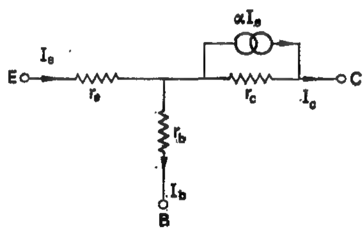


FIG. 9.1. A low-frequency equivalent T circuit for common base

The resistance of the base–collector junction is very much higher because this junction is reverse biased; we shall call this resistance r_c.

A *simplified* form of an equivalent electrical circuit is shown in Figure 9.1, where r_b represents the resistance of the base connection. This circuit is called a "common base low-frequency equivalent T circuit". Because of the Early effect (see 8.2.1) some modification of this basic circuit is necessary, but we will neglect it at this stage.

9.2. AMPLIFICATION IN COMMON BASE CONNECTION

If we consider the input circuit between the emitter E and the base B of Figure 9.1 then the input voltage between E and B is

$$V_{in} = I_e r_e + I_b r_b \tag{9.1}$$

by applying Kirchhoff's second law around the input loop.

But $I_b = I_e - I_c$ (the a.c. form of (7.1))

$= I_e - \alpha I_e$

Substituting for I_b in eqn. (9.1)

$$V_{in} = I_e r_e + (I_e - \alpha I_e) r_b$$

So

$$V_{in} = I_e(r_e + r_b(1-\alpha)) \tag{9.2}$$

If now we consider a load resistor of value R_L connected in the output circuit between C and B a current I_c will flow through it. If we call the voltage drop across this resistor V_{out} then

$$V_{out} = I_c R_L$$
$$= \alpha I_e R_L \tag{9.3}$$

The voltage gain is $V_{out}/V_{in} = A_V$, so by dividing eqn. (9.3) by (9.2) we have

$$A_V = \frac{\alpha I_e R_L}{I_e[r_e + r_b(1-\alpha)]}$$

$$= \frac{\alpha R_L}{re + (1-\alpha)r_b} \qquad (9.4)$$

Typical values might be $\alpha = 0{\cdot}98$, $R_L = 10\ \text{k}\Omega$, $r_e = 25\ \Omega$, $r_b = 250\ \Omega$; and substituting these values in (9.4) shows that

$$A_V = \frac{0{\cdot}98 \times 10{,}000}{25 + (1-0{\cdot}98)250} \simeq \frac{10^4}{30} \simeq 330.$$

As the current gain in common base connection is nearly unity, it is clear that the power gain achieved is about 330, or 26 db. (See Appendix 4, Vol. 1.)

The student should note that the example given represents "badly matched" conditions, and that one could expect a power gain of about 1500 in typical common base circuits.

The transistor will produce power gain in two other configurations: common emitter and common collector (analogous to the common cathode and common anode circuits in thermionic valve circuits—see Vol. 1, Chap. 6). The common emitter circuit is the most popular, and we will now consider it in detail.

9.3. A LOW-FREQUENCY EQUIVALENT CIRCUIT FOR COMMON EMITTER

If we take the equivalent circuit of Figure 9.1 and rearrange the geometry we obtain the common emitter low-frequency equivalent T circuit of Figure 9.2(a). The current through r_c of

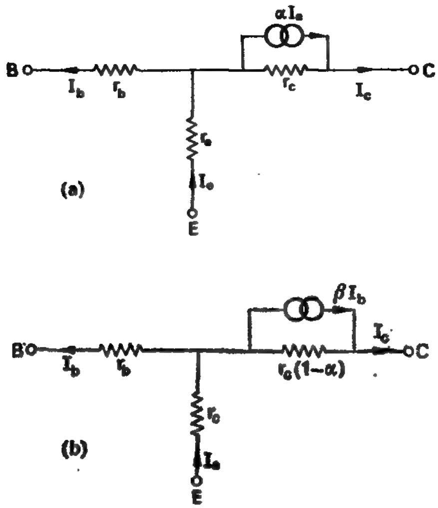


FIG. 9.2. Low-frequency equivalent T circuits for common emitter

$I_c-\alpha I_e$ causes a p.d. of V across it.

$$V = r_c(I_c-\alpha I_e)$$

But
$$I_e = I_b+I_c$$

$$\begin{aligned}\therefore V &= r_c(I_c-\alpha(I_b+I_c))\\ &= r_cI_c-\alpha r_cI_c-\alpha r_cI_b\\ &= r_c(1-\alpha)I_c-r_c\alpha I_b\\ &= r_c(1-\alpha)[I_c-\alpha I_b]/(1-\alpha)\\ &= r_c(1-\alpha)(I_c-\beta I_b)\end{aligned}$$

This voltage can be interpreted to mean a current of $I_c-\beta I_b$ flowing through a resistance of $r_c(1-\alpha)$. Figure 9.2(b) represents the new common emitter T equivalent circuit, wherein the current generator is now βI_b and the parallel resistance has been reduced from r_c to $r_c(1-\alpha)$, which is considerably less. (α is

usually very nearly 1.)

We had $\qquad I_c = \alpha I_e + I_{co}$ (A.C. form of $\qquad$ (7.6))

and $\qquad I_b = I_e - I_c \qquad$ (7.1)

Substituting for I_e from (7.1) into (7.6) gives us

$$I_c = \alpha(I_b + I_c) + I_{co};$$

$$I_c = \frac{\alpha}{1-\alpha} I_b + \frac{I_{co}}{1-\alpha} \tag{9.5}$$

A typical value for α is 0·98; then,

$$I_c = \frac{0{\cdot}98}{1-0{\cdot}98} I_b + \frac{I_{co}}{1-0{\cdot}98}$$

$$I_c = 49 I_b + 50 I_{co} \tag{9.6}$$

Equation (9.6) shows us that a change in base current I_b produces a corresponding change in collector current I_c some 50 times greater. The term $\alpha/(1-\alpha)$ is the current gain in common emitter configuration, and is often given the symbol β, or α' (see (a) in Section 8.2.2). In addition, eqn. (9.6) shows that when the base current I_b is zero, the collector to emitter leakage current is some fifty times the corresponding collector to base leakage current (I_{co}) in common base connection (see (b) in Section 8.2.2). The leakage current in common emitter, of value $I_{co}/(1-\alpha)$ is usually given the symbol I'_{co}. (Compare the output characteristics in common emitter, Fig. 8.18, with the corresponding ones for common base in Fig. 8.15.)

We saw in eqn. (2.8) that the electrical conductivity of semiconductor material is strongly temperature dependent, and so is I_{co} (the reverse saturation current in common base). As I'_{co} is very much larger it is clear that temperature stability plays a very important part in electronic circuits designed around transistors in common emitter configuration.

As the collector voltage is now the sum of the two junction voltages, even when the collector junction voltage is zero the emitter junction voltage remains and the "knee" of each I_B curve is displaced to the right as I_B increases (Figure 8.18; see also (c) in Section 8.2.2).

In order to explain (d) in Section 8.2.2, we must consider the small signal performance of the common emitter amplifier.

9.4. THE COMMON EMITTER AMPLIFIER

(see Expts. 9.1 and 9.2)

A simple amplifier, using a *p–n–p* transistor in common emitter configuration, is shown in Figure 9.3. The signal is supplied by a current generator equivalent circuit (a perfect current generator of i in parallel with the source resistance R_s (see Vol. 1, Sect. 6.8.1), and the amplified signal is fed to a pure resistance load R_L. As we are only interested in this section in the *signal* and its amplification, we omit the D.C. supplies and the steady values of currents and voltages, so no batteries are shown in Figure 9.3. (This can be justified by the superposition theorem, which states that "the current in any branch of a circuit is the algebraic sum of the currents that would be produced by each e.m.f. acting alone, all the other sources of e.m.f. being replaced meanwhile by their respective internal resistances".)

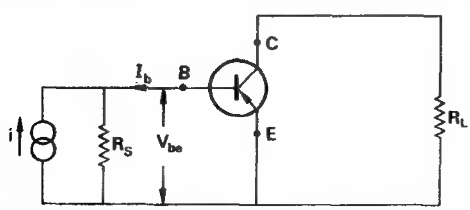


FIG. 9.3. A simple common emitter amplifier circuit

9.4.1. The input resistance, R_{in}

Let the signal current I_b be associated with a voltage V_{be}.

Then
$$R_{in} = \frac{V_{be}}{I_b} \tag{9.7}$$

But, from Figure 9.2(b), the voltage V_{be} is the sum of the p.d.s across r_b and r_e. So,

$$V_{be} = I_b r_b + I_e r_e \tag{9.8}$$

where I_e is the signal emitter current.

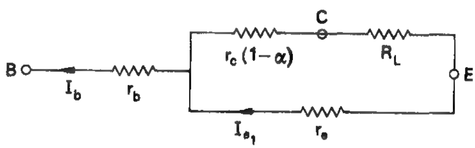


FIG. 9.4. Calculation of input resistance

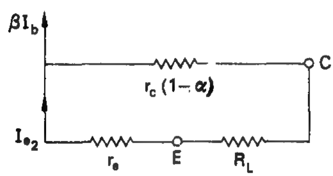


FIG. 9.5. Calculation of input resistance

To find the current I_e we use the Superposition Theorem; first, consider the current I_{e_1} in r_e due to I_b *alone* (see Fig. 9.4)

$$I_{e_1} = I_b \times \frac{r_c(1-\alpha)+R_L}{r_e+r_c(1-\alpha)+R_L} \tag{9.9}$$

Secondly, consider the current I_{e_2} in r_e due to βI_b *alone* (see Fig. 9.5)

$$I_{e_2} = \beta I_b \times \frac{r_c(1-\alpha)}{r_e+r_c(1-\alpha)+R_L} \tag{9.10}$$

So
$$I_e = I_{e_1}+I_{e_2}$$
$$= I_b\left[\frac{r_c(1-\alpha)+R_L+\beta r_c(1-\alpha)}{r_e+r_c(1-\alpha)+R_L}\right] \tag{9.11}$$

Substituting for I_e from (9.11) into (9.8) we have

$$V_{be} = I_b r_b+I_b r_e\left[\frac{(1+\beta)(1-\alpha)r_c+R_L}{r_e+r_c(1-\alpha)+R_L}\right] \tag{9.12}$$

Substituting for V_{be} from (9.12) into (9.7) we have

$$R_{\text{in}} = r_b+r_e\left[\frac{(1+\beta)(1-\alpha)r_c+R_L}{r_e+r_c(1-\alpha)+R_L}\right] \tag{9.13}$$

Now let $r_c(1-\alpha) = r_d$, and let $1+\beta+R_L/r_d$ tend to β, and neglect r_e/r_d, which is approximately 10^{-3}. Then,

$$R_{\text{in}} \simeq r_b+\beta\frac{r_e}{1+(R_L/r_d)} \tag{9.14}$$

This is our expression for the input resistance to the transistor. Note that it depends on the output circuit, because of the term R_L/r_d. Typical values would be: $r_b = 1\ \text{k}\Omega$, $r_e = 10\ \Omega$, $r_d = 20\ \text{k}\Omega$, $\beta = 100$, and the load R_L might be 10 kΩ. Substitution in (9.14) shows that $R_{\text{in}} = 1670\ \Omega$, which is typical of the order of magnitude for the input resistance to an average transistor in common emitter configuration.

9.4.2. The current amplification, A_I

The current amplification produced by the transistor is the ratio of the change in current I_e in the load to the corresponding change in current I_b at the base. To find I_e we must again use the superposition theorem: first, consider the current I_{e_1} in R_L

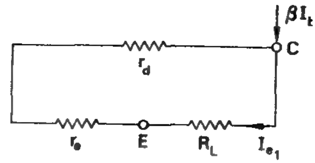


FIG. 9.6. Calculation of current amplification

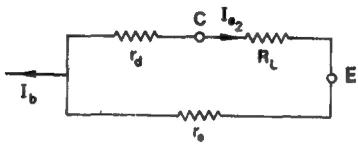


FIG. 9.7. Calculation of current amplification

due to βI_b *alone* (see Fig. 9.6):

$$I_{e_1} = \beta I_b \times \frac{r_d}{r_e + R_L + r_d} \tag{9.15}$$

Secondly, consider the current I_{e_2} in R_L due to I_b *alone* (see Fig. 9.7):

$$I_{e_2} = -I_b \times \frac{r_e}{r_e + R_L + r_d} \tag{9.16}$$

So
$$I_e = I_{e_1} + I_{e_2}$$
$$= \frac{\beta I_b r_d - I_b r_e}{r_e + R_L + r_d}$$

but:
$$A_I = \frac{I_e}{I_b}$$
$$= \frac{\beta r_d - r_e}{r_e + R_L + r_d}$$
$$= \frac{\beta - (r_e/r_d)}{1 + (R_L/r_d) + (r_e/r_d)}$$

Now r_e/r_d is typically about 10^{-3}, so

$$A_I \simeq \frac{\beta}{1+R_L/r_d} \tag{9.17}$$

This is our expression for the current gain. Note that in the ideal case the maximum gain is β where $\beta = I_c/I_b$, but in practice it is less than this because r_d is never infinite. Notice the analogy to the thermionic valve voltage gain of

$$\frac{\mu R}{R+r_a} = \frac{\mu}{1+(r_a/R)}$$

where the maximum gain possible is the amplification factor μ, but the gain in practice is always less than this because of the finite value of r_a.

For the transistor, typical values are $\beta = 100$, $R_L = 10$ kΩ, and $r_d = 20$ kΩ, giving a current gain of about 67.

9.4.3. The voltage amplification, A_V

The change in base current I_b associated with the input resistance R_{in} produces a change of input voltage V_{in}. Similarly the associated change in load current I_e produces a change in output voltage of V_{out}, where $V_{out} = I_e \times R_L$. The voltage amplification is V_{out}/V_{in}, so

$$A_V = \frac{-I_e R_L}{I_b R_{in}} = -A_I \frac{R_L}{R_{in}}$$

$$= -\frac{\beta}{1+(R_L/r_d)} \times \frac{R_L}{r_b + \dfrac{\beta r_e}{1+(R_L/r_d)}}$$

$$\therefore A_V = \frac{-\beta R_L}{r_b[1+(R_L/r_d)]+\beta r_e} \tag{9.18}$$

As an example, if $\beta = 100$, $R_L = 10$ kΩ, $r_b = 1$ kΩ, $r_e = 10$ Ω, $r_d = 20$ kΩ, the voltage gain is about 400.

9.4.4. The power amplification, A_P

The power amplification is just the product of the current and voltage amplifications, so,

$$A_P = A_I A_V \tag{9.19}$$

Using the previous information, if $A_I = 67$ and $A_V = 400$, then $A_P = 27{,}000$, or 43 db.

9.4.5. The output resistance, R_{out}

We can see from Figure 9.2(b) that the output resistance seen between the terminals C and E is

$$R_{out} = r_c(1-\alpha)+r_e \tag{9.20}$$

A more accurate assessment using Thenevin's theorem shows that R_{out} is given by

$$R_{out} = r_c(1-\alpha)+\frac{\alpha r_c r_e}{R_s+r_b+r_e} \tag{9.21}$$

Note that R_{out} is dependent on the source resistance, just as input resistance R_{in} is dependent on the load R_L (eqn. (9.14)).

Typical values of $\alpha = 0{\cdot}99$, $r_c = 2\ \mathrm{M}\Omega$, $r_b = 1\ \mathrm{k}\Omega$, $R_s = 1\ \mathrm{k}\Omega$, and $r_e = 10\ \Omega$, give an output resistance of nearly 30 kΩ.

9.5. DETERMINATION OF PARAMETERS FROM TRANSISTOR CHARACTERISTICS

The equivalent T circuit parameters r_e, r_b, and r_c, are not determinable purely from the input or output (collector) characteristics of a transistor alone; this is unfortunate, because the T equivalent circuit bears the closest resemblance to the physical

construction of the transistor. However, there are several other electrical parameters, such as y, z, and h; and h can be found directly from the characteristics. The justification for these parameters (the h-parameters) can be found in Appendix I (see Expt. 9.3).

The input characteristics of a typical small power, low-frequency transistor (with typical modesty we will call it the JJ1) are represented in Figure 9.8. Consider a base current of 40 μA and the $V_{CE} = -15$ V curve. The reciprocal of the slope of the

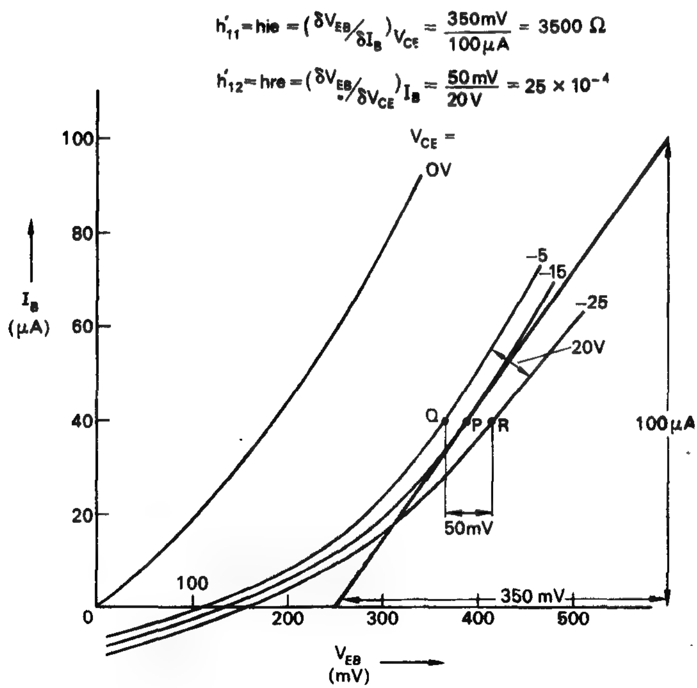


FIG. 9.8. Input characteristics for a JJ1 transistor

curve at the point (shown as the tangent at point P) is

$$350 \text{ mV}/100 \ \mu\text{A} = 3500 \ \Omega.$$

This is the parameter called h'_{11} (or h_{ie}). In addition, the value of V_{EB} increases from 365 to 415 mV as we go from Q to R (with a constant I_B of 40 μA) but V_{CE} goes from -5 to -25 V. The ratio of 50 mV/20 V $= 25\times10^{-4}$ is labelled h'_{12} (or h_{re}).

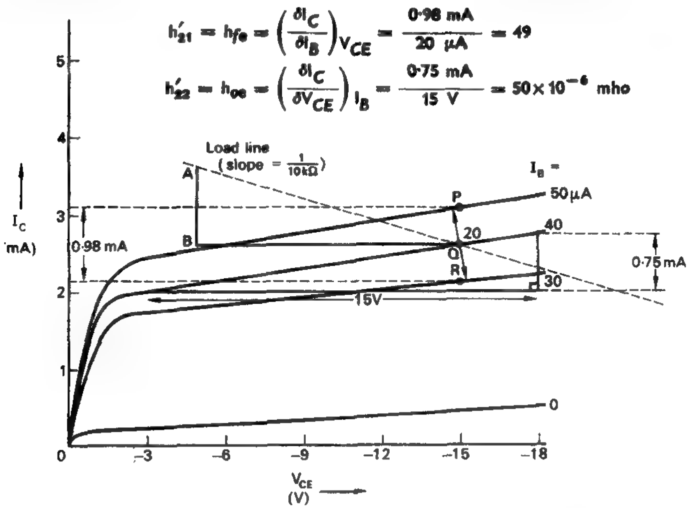


FIG. 9.9. Output characteristics for a JJ1 transistor

The output (or collector) characteristics of the JJ1 transistor are shown in Figure 9.9. As we go from R to P (at a constant value of $V_{CE} = -15$ V), I_C increases from 2·10 to 3·08 mA, but the base current I_B increases from 30 to 50 μA. The ratio 0·98 mA/20 μA $= 49$ is h'_{21} (or h_{fe}).

Finally, the slope of the $I_B = 40$ μA line at Q is 0·75 mA/15 V $= 50\times10^{-6}$ mho, called h'_{22} (or h_{oe}).

It is most important for the student to realize that these h values are *small signal parameters*; they are *not* constants, even

for a given transistor. In the above examples they have been calculated for $V_{CE} = -15$ V, $I_B = 40\ \mu$A. If they were determined anywhere else on the same characteristics, they might well have different values.

In Appendix II the relationships between the h parameters and the T parameters are derived for the common emitter configuration. Here we will assume these relationships, and, by substituting our determined values for the h-parameters, we will find the corresponding r_e, r_b, r_c, and α values.

Substituting the values

$$h'_{11} = h_{ie} = 3500 \text{ ohm},$$
$$h'_{12} = h_{re} = 25\times 10^{-4},$$
$$h'_{21} = h_{fe} = 49,$$
$$h'_{22} = h_{oe} = 50\times 10^{-6} \text{ mho},$$

we find that

$$r_e = \frac{h_{re}}{h_{oe}} = \frac{25\times 10^{-4}}{50\times 10^{-6}\text{ mho}} = 50\ \Omega;$$

$$r_b = h_{ie} - \frac{h_{re}(1+h_{fe})}{h_{oe}}$$

$$= 3500 - \frac{25\times 10^{-4}(1+49)}{50\times 10^{-6}}$$

$$= 3500 - 2500 = 1000\ \Omega;$$

$$r_c = \frac{1+h_{fe}}{h_{oe}}$$

$$= \frac{1+49}{50\times 10^{-6}} = 1\ \mathrm{M}\Omega;$$

$$\alpha = \frac{h_{fe}+h_{re}}{1+h_{fe}}$$

$$= \frac{49+25\times 10^{-4}}{1+49} = 0{\cdot}98;$$

$$r_d = r_c(1-\alpha)$$
$$= 10^6\,(1-0{\cdot}98) = 20\ \mathrm{k}\Omega,$$

and $$\beta = \frac{\alpha}{1-\alpha} = \frac{0{\cdot}98}{0{\cdot}02} = 49$$

So the JJ1 transistor in common emitter configuration can be thought of (at low frequencies) as the electrical circuit of Figure 9.10.

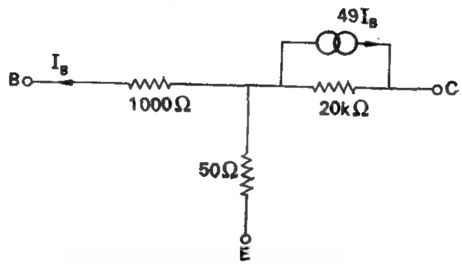


FIG. 9.10. The low-frequency equivalent *T* circuit for the JJ1 transistor in common emitter

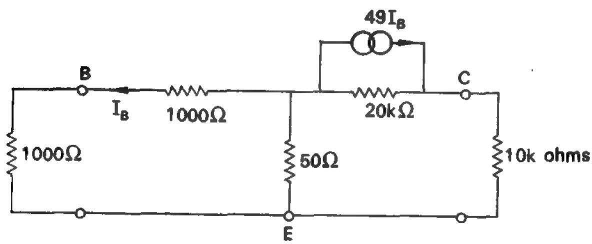


FIG. 9.11. The circuit of Figure 9.10 feeding a 10 kΩ load from a 1 kΩ source

Let us now predict the performance of the JJ1 transistor as an amplifier in common emitter configuration. Figure 9.11 shows the *T* equivalent circuit with the values already determined being fed from a source of 1000 Ω and feeding into a load of 10,000 Ω.

Using the exact form of the expression for the input resistance

R_{in} (eqn. (9.13)) we have

$$R_{in} = r_b + r_e \left[\frac{(1+\beta)(1-\alpha)r_c + R_L}{r_e + r_c(1-\alpha) + R_L} \right] \tag{9.13}$$

giving
$$R_{in} = 1000 + 50 \left[\frac{(1+49)(1-0{\cdot}98)10^6 + 10^4}{50 + 10^6(0{\cdot}02) + 10^4} \right]$$
$$= 1000 + 1680$$
$$= 2680\ \Omega$$

Using the approximate form of the expression for the current gain A_I we have:

$$A_I = \frac{\beta}{1+(R_L/r_d)} \tag{9.17}$$
$$= \frac{49}{1+(10^4)/(2\times 10^4)} = \frac{49}{3/2}$$
$$= \mathbf{32{\cdot}7}$$

Using the expression for the voltage gain we have

$$A_V = \frac{-\beta R_L}{r_b(1+(R_L/r_d)) + \beta r_e}$$
$$= \frac{-49\times 10^4}{1000(1+10^4/(2\times 10^4)) + 49\times 50}$$
$$= \mathbf{-124}$$

The power gain is

$$A_P = A_I A_V$$
$$= 32{\cdot}7\times 124$$
$$= 4060$$

Using the approximate form of the expression for the output

resistance R_{out} we obtain

$$R_{out} = r_c(1-\alpha)+\frac{\alpha r_c r_e}{R_s+r_b+r_e}$$

$$= 10^6\,(1-0{\cdot}98)+\frac{0{\cdot}98\times10^6\times50}{10^3+10^3+50}$$

$$= 20\ \text{k}\Omega+24\ \text{k}\Omega$$

$$= 44\ \text{k}\Omega$$

Of course, the input and output resistances and gains can be expressed directly in h-parameters, yielding precisely the same answers (see Appendix III).

9.6. THE LOAD LINE

Figure 9.12 represents the conventional circuit of the JJ1 transistor fed from a 1 kΩ source and feeding a 10 kΩ load.

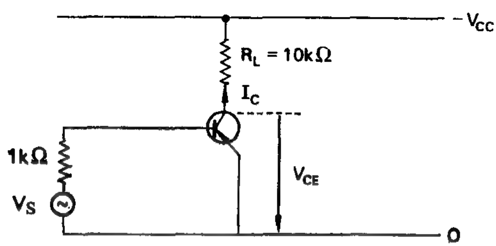


FIG. 9.12. The conventional circuit of Figure 9.11

By Kirchhoff's second law,

$$-V_{CC} = V_{CE}-I_C R_L$$

or
$$+V_{CC} = I_c R_L-V_{CE}$$

So
$$I_C = \frac{1}{R_L}\cdot V_{CE}+\frac{V_{CC}}{R_L}$$

This represents a straight line on the I_C/V_{CE} graph (Fig. 9.9) of slope $-1/R_L$, where R_L is the load resistance. This straight line is called the "load line".

The load line appropriate to a load of 10 kΩ is shown by the dotted line on the output characteristics of the JJ1 transistor in Figure 9.9. The reader can check this by noting that $AB = 3{\cdot}6-2{\cdot}6 = 1$ mA, and $BQ = -(15-5) = -10$ V,

and $$\frac{AB}{BQ} = \frac{1\text{ mA}}{10\text{ V}} = \frac{1}{10\text{ k}\Omega}.$$

The load line has been constructed to pass through the point $I_B = 40$ μA, and $V_{CE} = -15$ V, because this was the point at which the incremental parameters were found: we have now made Q the "operating" or "quiescent" point of the JJ1 transistor amplifier.

The current gain is easily found from the interception of the load line with the $I_B = 30$ and 50 μA curves. The corresponding values of I_C for these points are 2·20 and 2·85 mA, so the gain of the amplifier

$$= \frac{\Delta I_C}{\Delta I_B} = \frac{(2{\cdot}85-2{\cdot}20)10^{-3}}{(50-30)10^{-6}}$$

$$= \frac{0{\cdot}65}{20}\times 10^3$$

$$= \frac{65}{2} \doteq 33.$$

This is the same answer as found from the h-parameters (see Sect. 9.5).

The JJ1 transistor has to be "biased" at the point Q, where $I_B = 40$ μA and $V_{CE} = -15$ V by the application of a *direct* current of 40 μA to the base of the transistor: the *alternating* current signal applied to the base then causes variation of the steady 40 μA base current, creating a variation of the same fre-

quency in the collector current, which is far larger than the original signal applied to the base. This is what is called current *amplification.* We must now consider how a steady direct current can be applied to the base of the transistor.

9.7. BIAS CIRCUITS

It is not a simple matter of applying the right base current I_B to the transistor and operating it with the correct voltage V_{CE} applied across it. This would certainly "bias" it, i.e. would set the operating and quiescent points, but it would not make any allowance for the input signal to vary the currents a small amount either way, nor would it take any account of current changes produced by temperature changes. And there are bound to be temperature changes, because a current I flowing in a transistor circuit of effective resistance R produces heat at the rate I^2R, just as in any other kind of resistance. If we can also make allowance for changes due to ageing or changing of components, so much the better. So stabilization of the operating point will not be so easy for the transistor as for (say) the vacuum triode. In practice, considerable variation exists between transistors of the same type number; for example, Mullard quote the value of h'_{fe} for an OC36 as between 30 and 110.

As usual with an amplifying device the operating point is specified by reference to the size of the input signal, the load, and the distortion level which can be permitted at the output. With transistors, temperature changes cause large variations in leakage currents, but fortunately, in designing circuits to minimize this effect, we usually also cope with the problem of ageing or changing components.

We shall again only consider the common emitter configuration, as it is by far the commonest.

9.7.1. Fixed bias

In order to inject a known current into the base it would seem obvious to connect a bias resistor R_B between the base and the negative supply line, as in Figure 9.13.

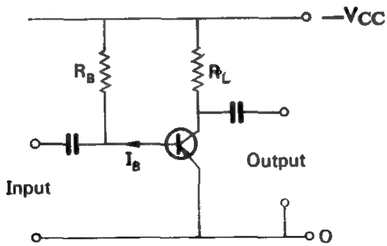


FIG. 9.13. The circuit for fixed bias

If we neglect the voltage across the emitter–base junction, the *PD* across R_B is approximately V_{CC}, so

$$R_B = \frac{V_{CC}}{I_B}.$$

However, this simple base current stabilization circuit is unsatisfactory for most purposes, because the circuit does not compensate for temperature and component variations; nor is the collector–emitter leakage current controlled.

9.7.2. Self bias

This method is shown in Figure 9.14 and is simply the circuit of Figure 9.13 with one end of R_B reconnected, producing collector voltage feedback, giving a measure of stabilization. Any increase in collector–emitter leakage current will cause V_{CE} to fall and hence I_B will fall, and so will the effective collector cur-

rent. In practice R_L is much less than R_B, so R_B is still approximately given by V_{CE}/I_B. However, as there *is* some dependence on R_L, the degree of control depends on the load.

A disadvantage of this simple circuit is that A.C. negative feedback also occurs, reducing the stage gain; but this can be

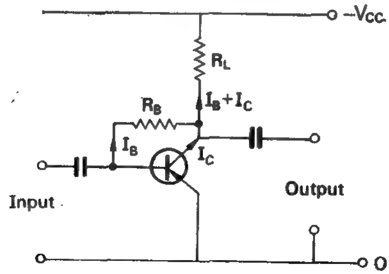


FIG. 9.14. The circuit for self bias

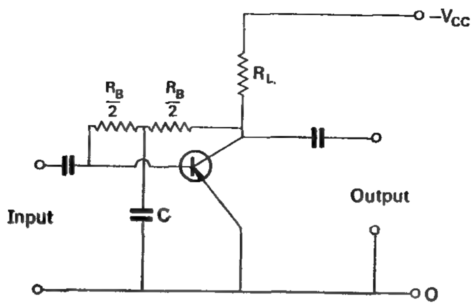


FIG. 9.15. An alternative circuit for self bias

overcome by splitting the bias resistor R_B into two equal halves and decoupling the mid-point by a capacitor C (Fig. 9.15). However, this loads both the input and the output. (Note that "decoupling" earths the A.C. voltage.)

9.7.3. Stabilized bias

The stability of the previous circuit can be greatly improved by inserting a resistor R_3 in the emitter lead, which then controls emitter current variation. But this resistor produces a *reverse* bias at the base-emitter junction, and a potentiometer consisting of R_1 and R_2 is added to correct this (Fig. 9.16).

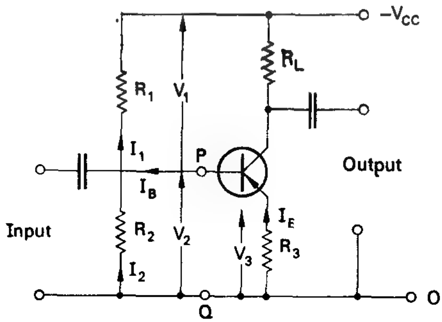


FIG. 9.16. The base potentiometer stabilized bias circuit

The base current I_B is now dependent on $-(V_2-V_3) = V_{BE}$. If I_E should increase for any reason, V_{BE} will decrease, and so will I_B. As I_B flows through R_1, V_1 will decrease and so V_2 will increase, because $V_1+V_2 = V_{CC}$, which is constant. As V_2 increases it offsets the fall in V_{BE}, thus stabilizing the operating point. Clearly R_1 and R_2 should be as small as possible and then I_1 will be much greater than I_B, and V_2 will be nearly constant. However, we have the conflicting points that if R_1 and R_2 are small, the constant drain current taken from the supply voltage V_{CC} will be large, and also the signal input will be loaded; so a compromise must be reached.

We will take the analysis of this most important circuit a little further. The circuitry to the left of the points P and Q in Fig-

ure 9.16 can be thought of as a voltage source V, where

$$V = \frac{V_{CC}R_2}{R_1+R_2} = V_2$$

in series with a resistance R, where

$$R = \frac{R_1R_2}{R_1+R_2}$$

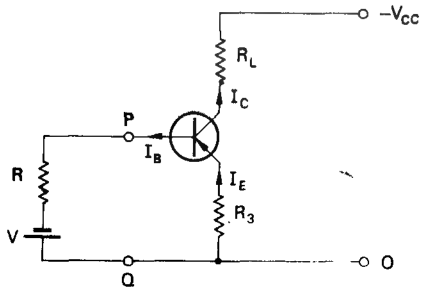


FIG. 9.17. The Thevenin equivalent of the circuit in Figure 9.16

This follows from Thevenin's theorem. We now have the circuit of Figure 9.17. Applying Kirchhoff's second law around the left-hand loop, we have

$$I_ER_3+V_{BE}+I_BR = V \tag{9.22}$$

where $V_{BE}=PD$ across the base–emitter junction, and is assumed constant.

Applying Kirchhoff's first law, we have

$$I_E = I_C+I_B \tag{9.23}$$

Also,
$$I_c = \frac{\alpha}{1-\alpha}I_b+\frac{I_{co}}{1-\alpha} \tag{9.5}$$

From 9.5,
$$I_b = \frac{1-\alpha}{\alpha}I_c-\frac{I_{co}}{\alpha} \tag{9.24}$$

Substituting for I_B, I_E from (9.24) and (9.23) into (9.22), we have

$$(I_C+I_B)R_3+V_{BE}+\frac{1-\alpha}{\alpha}I_CR-\frac{I_{co}R}{\alpha}=V$$

$$\therefore \left[I_C+\frac{1-\alpha}{\alpha}I_C-\frac{I_{co}}{\alpha}\right]R_3+\frac{1-\alpha}{\alpha}I_CR=\frac{I_{co}R}{\alpha}+(V-V_{BE})$$

$$\therefore \alpha I_CR_3+(1-\alpha)I_CR_3+(1-\alpha)I_CR=I_{co}(R+R_3)+\alpha(V-V_{BE})$$

$$\therefore I_CR_3+(1-\alpha)I_CR=I_{co}(R+R_3)+\alpha(V-V_{BE})$$

$$\therefore I_C[R_3+(1-\alpha)R]=I_{co}(R+R_3)+\alpha(V-V_{BE}) \tag{9.25}$$

The ratio $\delta I_C/\delta I_{co}=S$ is called the stability factor, and is a measure of the change of I_C due to a change of I_{CO} caused, perhaps, by a rise in the temperature of the transistor. Clearly we need S as small as possible; a value less than 10 is usually sufficient for most purposes. From eqn. (9.25),

$$S=\frac{\delta I_C}{\delta I_{CO}}=\frac{R+R_3}{R_3+(1-\alpha)R} \tag{9.26}$$

9.8. DESIGN OF A BIAS CIRCUIT

(a) A D.C. load line is first chosen, and a suitable operating point Q on the output characteristics of the given transistor is specified. Q is chosen so that the transistor is not overloaded (see Fig. 9.18), distortion is kept to a minimum, and the A.C. requirements (such as current gain) are satisfied. Because we now have an emitter resistor R_3 the intercept of the load line on the I_C axis is now

$$\frac{V_{CC}}{R_L+R_3}$$

and not V_{CC}/R_L as in (9.6). So V_{CC}, R_L, and R_3 have now been determined.

(b) Now we make the rule-of-thumb assumption that the volts lost on emitter R_3 should not exceed a quarter of the battery voltage, i.e.

$$I_E R_3 \simeq \frac{V_{CC}}{4}, \quad \text{or} \quad 10V_{BE}.$$

This should give a value for R_3 of between 500 and 2000 Ω, typical for small transistors.

(c) Choose R_2 to lie in value between $5R_3$ and $10R_3$. This usually produces adequate stability without overloading the input signal.

(d) To find R_1 we must remember that it has to give the right value of I_B. We had

$$\begin{aligned} V_2 &= V_3 + V_{BE} \\ &= I_C R_3 + V_{BE}, \text{ assuming that } I_E \simeq I_C. \end{aligned}$$

Also,

$$\begin{aligned} V_1 &= I_1 R_1 \\ &= (I_2 + I_B)R_1 \\ &= \left(\frac{V_2}{R_2} + I_B\right)R_1 \\ &= \left(\frac{I_C R_3 + V_{BE}}{R_2} + I_B\right)R_1 \end{aligned}$$

Now

$$\begin{aligned} V_{CC} &= V_1 + V_2 \\ &= \left(\frac{I_C R_3 + V_{BE}}{R_2} + I_B\right)R_1 + I_C R_3 + V_{BE} \end{aligned}$$

So

$$R_1 = \frac{V_{CC} - (I_C R_3 + V_{BE})}{[(I_C R_3 + V_{BE})/R_2] + I_B}$$

From the point of view of this approximate evaluation, the value of V_{BE} can be taken as 0·2 or 0·6 V, depending on whether the transistor is germanium (e.g. *A*CY19) or silicon (e.g. *B*FY10)

respectively. (Under a recent classification system, the first letter identifies the material of any particular transistor type. See Appendix IV.)

(e) Check that the stability factor is less than 10.

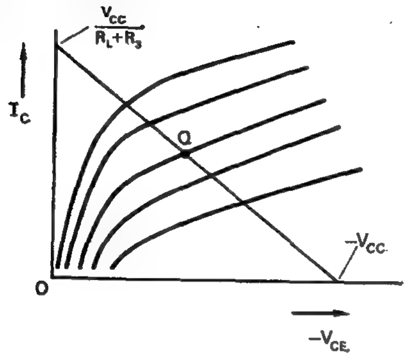


FIG. 9.18. Design of a bias circuit

The following example should make clear the design of a common emitter amplifier, with base potential divider stabilization.

Figure 9.19 shows the output characteristics of the JJ1 transistor (check with Fig. 9.9). We assume it is a germanium type, and $V_{BE} = 0{\cdot}2$ V.

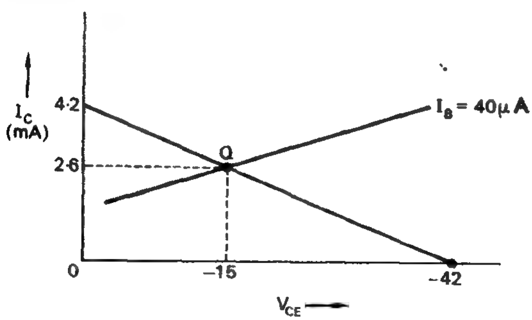


FIG. 9.19. Representation of the output characteristic of the JJ1 transistor

(a) If $V_{CC} = -42$ V, and $V_{BE} = 0{\cdot}2$ V, our load line (i.e. $R_L + R_3$) intercepts the I_C axis at 4·2 mA.

$$\text{So } R_L + R_3 = \frac{42}{4{\cdot}2 \times 10^{-3}} = 10 \text{ k}\Omega.$$

(b) $\dfrac{V_{CC}}{4}$ 10 V, and $10V_{BE} = 2$V.

Let $I_E R_3 = 2{\cdot}6$ V, and let I_E at the point Q be 2·6 mA.

$$\text{So } R_3 = \frac{2{\cdot}6 \text{ V}}{2{\cdot}6 \text{ mA}} = 1 \text{ k}\Omega.$$

So R_L must be 10 kΩ $-$ 1 kΩ $=$ 9 kΩ.

(c) $5R_3 = 5$ kΩ, and $10R_3 = 10$ kΩ. So let R_2 be 7 kΩ.

$$\text{(d) } R_1 = \frac{V_{CC} - (I_C R_3 + V_{BE})}{[(I_C R_3 + V_{BE})/R_2] + I_B}$$

$$= \frac{42 - (2{\cdot}6 \times 10^{-3} \times 10^3 + 0{\cdot}2)}{[(2{\cdot}6 \times 10^{-3} \times 10^{+3} + 0{\cdot}2)/7 \times 10^3] + 40 \times 10^{-6}}$$

$$\simeq 89 \text{ k}\Omega.$$

$$\text{(e) } S = \frac{R + R_3}{R_3 + (1-\alpha)R} \quad \text{where} \quad R = \frac{R_1 R_2}{R_1 + R_2} = \frac{89 \times 7}{89 + 7}$$

$$\therefore \; R = 6{\cdot}5 \text{ k}\Omega$$

$$S = \frac{6{\cdot}5k + 1k}{1k + (1 - 0{\cdot}98)6{\cdot}5k} \text{ because } \alpha = 0{\cdot}98$$

$$S = \frac{7{\cdot}5}{1{\cdot}13} \text{ which is less than 10, and satisfactory.}$$

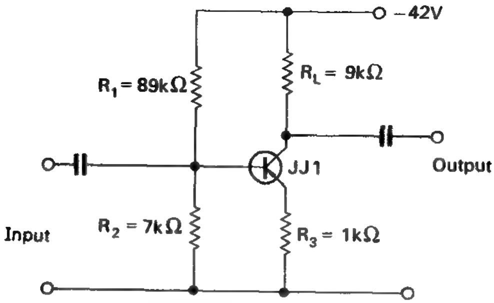


FIG. 9.20. The JJ1 transistor in use as a common emitter audio-frequency amplifier

The component values so found are shown in Figure 9.20, which shows the JJ1 transistor in use as a common emitter A.F. amplifier.

9.9. ALTERNATIVE TRANSISTOR CONFIGURATIONS

The reader will recall that we began this chapter by considering the transistor in common base configuration. We then concerned ourselves solely with the common emitter configuration, because it is by far the commonest. An alternative and fairly common way of using a transistor is in common collector configuration, wherein a signal is applied to the base and the output taken from the emitter. Here we will simply compare the features of the three modes of connection, in Table 9.1.

EXPERIMENT 9.1: Basic transistor current amplification (common emitter).

Apparatus required: D.C. supply, about 4·5 V at 10 mA.
Meter, centre-zero, 1–0–1 mA.
Meter, 0–10 mA.

Resistor, 47 kΩ, 1/4 W.
Transistor type OC81.
V.L.F. A.C. generator, Nuffield type ("White").
2 V accumulator.

TABLE 9.1

	Input impedance	Output impedance	Current gain	Voltage gain
Common emitter	Reasonably low (1 kΩ)	Reasonably high(20 kΩ)	Large (20→200)	Large
Common base	Very low (50 Ω)	Very high (200 kΩ)	Less than unity	Large
Common collector	Very high (1 MΩ)	Very low (50 Ω)	Large	Less than unity

Procedure. Connect up the circuit of Figure 9.21, and turn the VLF generator very slowly. Note the swing of the input current meter (typically 0·08 to 0·16 mA), and the much larger swing of the output current meter (typically 5·5 to 9·5 mA).

Calculate the "current gain", $\left[\frac{\Delta I_{out}}{\Delta I_{in}}\right]$

EXPERIMENT 9.2: The transistor as a voltage amplifier.

Apparatus required:

A.F. oscillator, having output of about 2 V peak to peak.
Capacitor, 1 μF, 12 V; (watch out for "electrolytic" capacitors, which must be connected the right way round).
Resistors, 1/4 W, 47 kΩ and 2 kΩ.
Transistor type OC81.
D.C. supply, 9 V, 20 mA.
Oscilloscope, e.g. Serviscope Minor.

Procedure. Connect up the circuit of Figure 9.22 and switch on the supply and A.F. oscillator. Connect the earth terminal of the oscilloscope to the

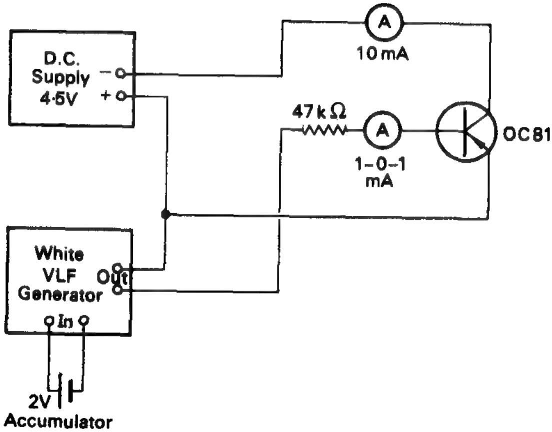


FIG. 9.21.

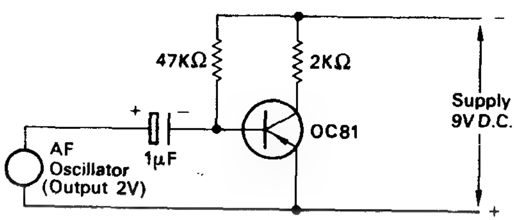


FIG. 9.22.

+ side of the supply, and the input terminal alternately to points *A* and *B*. Use the oscilloscope to measure the voltage gain (alternating voltage at *B* ÷ alternating voltage at *A*). Adjust the A.F. oscillator frequency and find at what frequencies the voltage gain falls to about half its maximum value, which should be at around 1 kHz.

Precaution. In all transistor circuits, malfunctioning may be caused if the D.C. supply has a high resistance. It can usually be cured with a large capacitor, say 1000 μF, connected across the supply.

EXPERIMENT 9.3: The h-parameters (see Appendix I)

Apparatus required:

1 transistor type OC81.
1 resistor, 100 kΩ.
1 resistor, 10 kΩ.
1 resistor, 100 Ω.
1 inductance, 10 henries or more.
1 capacitor, 500 μF or more.
1 A.F. generator with low *and* high impedance outputs (e.g. 3 and 600 Ω).
1 valve voltmeter with scales capable of reading 1 mV and 1 V. Supply, 5 V 20 mA.

Procedure. (a) To find h'_{11}, set up the circuit of Figure 9.23(a). Connect the valve voltmeter, set to A.C. (or with a 1 μF capacitor in series with the input) across the 100 Ω resistance. The varying component of I_B (I_b) is found by dividing the voltage reading (in volts) by 100; or to get I_b in mA, multiply the voltage reading by 10.

The varying component of V_{EB} is read directly by connecting the valve voltmeter between emitter and base. Suitable A.F. oscillator settings are shown on the diagram.

(b) To find h'_{12}, use the circuit of Figure 9.23(b). The varying components of V_{EB} and V_{CE} (∂V_{EB}, ∂V_{CE}) are read directly when the valve voltmeter is connected between emitter and base, and collector and emitter, respectively.

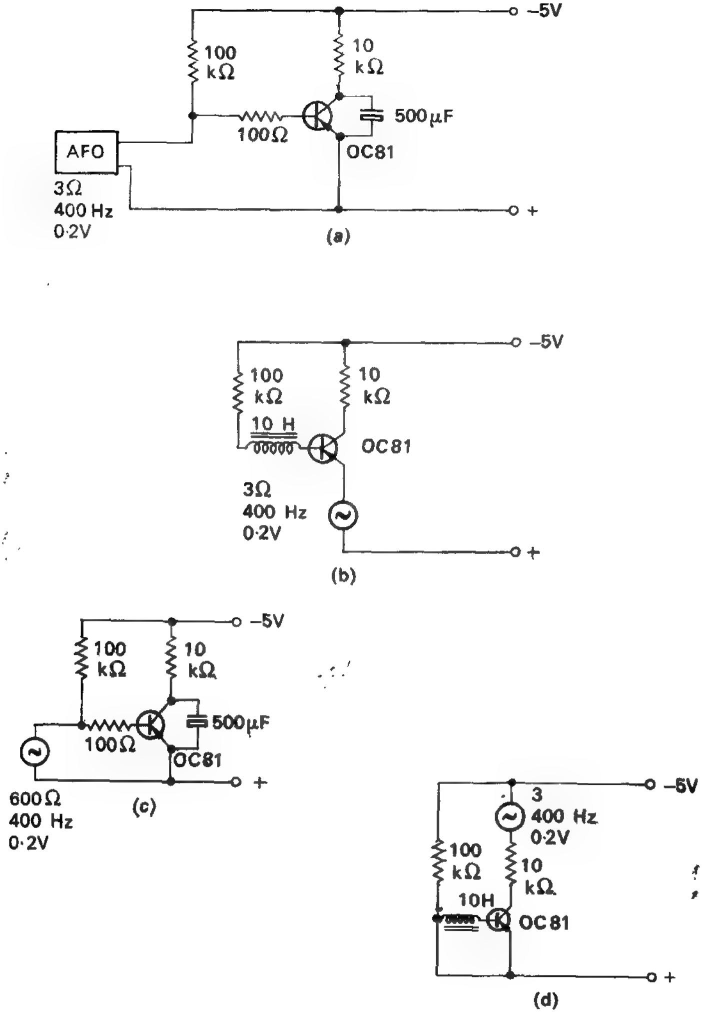
(c) To find h'_{21}, use the circuit of Figure 9.23(c). I_b is found (in mA) by connecting the vvm across the 100 Ω resistance, and multiplying the voltage reading by 10. I_c is found by reading the voltage across the 10 kΩ register, and *dividing* by 10 (if the answer is required in mA).

(d) To find h'_{22}, the circuit of Figure 9.23(d) is used. I_c is found in mA by reading the voltage across the 10 kΩ resistor, and dividing by 10. V_{ce} is read directly when the vvm is connected across collector and emitter.

Further study. Compare the values obtained in this experiment with those found in Expt. 8.2, using the relationships given in Appendix II.

Find these parameters at various frequencies, e.g. 10, 400, 10^4 Hz.

If a transistor such as the OC45 is substituted, interesting results may be obtained at frequencies of several MHz, but the inductance and capacitance will both need to be reduced, and inter-lead capacitances will begin to affect the readings. Also the vvm may not be reliable at radio frequencies.



–5V
100
kΩ
10
kΩ
100Ω
500μF
AFO
OC81
3Ω
400 Hz
0·2V
+
(a)
–5V
100
kΩ
10
kΩ
10 H
OC81
3Ω
400 Hz
0·2V
+
(b)
–5V
100
kΩ
10
kΩ
500μF
100Ω
OC81
+
600Ω
400 Hz
0·2V
(c)
–5V
3
400 Hz
0·2V
100
kΩ
10
kΩ
10H
OC81
+
(d)

FIG. 9.23.

QUESTIONS

1. Find an approximate value for the power gain of a transistor in common base configuration, using the following data:

$$\alpha = 0{\cdot}99; \quad R_L = 5\ \text{k}\Omega; \quad r_e = 15\ \Omega; \quad r_b = 330\ \Omega.$$

What would be the (alternating component of the) output current for an input current of 5 μA?

2. A common emitter amplifier has a base resistance of 1000 Ω, an emitter resistance of 22 Ω, a load resistance of 470 Ω, and a transistor having $\beta = 150$ and a base-collector junction resistance of 1 MΩ. Find the input resistance of the stage, and the current it would draw from a magnetic pick-up generating 15 mV in a source resistance of 4000 Ω.

3. Using the data of question 2, determine approximately the output impedance, output current, and current gain.

4. With the aid of a diagram, show what factors you would consider in choosing an "operating" or quiescent point on the output characteristics of a given transistor.

Give a circuit diagram showing how you would bias the transistor so as to work with this chosen operating point.

Briefly say what would happen if the input signal were to cause the transistor to work beyond the linear part of its output characteristic.

5. Define the hybrid parameters of a transistor in common emitter configuration and derive expressions giving their relationship to the current gain and incremental resistances.

(C. and G., Adv. Telecomm. and Electronic Principles, I, May 1967)

6. (i) Draw and explain a circuit diagram showing the basic form of a common emitter transistor amplifier.

(ii) Explain why: (a) the voltage gain of a transistor amplifier depends upon the internal resistance of the signal source; (b) the input impedance of a transistor amplifier is affected by the load impedance with which it operates; (c) there is a power gain between input and output circuits.

(U.L.C.I., Electrical Technology 1B, May 1966)

7. A common emitter amplifier uses a transistor having the following small-signal parameters: $\alpha = 0{\cdot}95$; $r_b = 500\ \Omega$; $r_e = 25\ \Omega$; $r_c = 500\ \text{k}\Omega$.

The load resistor is 10 kΩ. Calculate the input resistance and the voltage gain of this amplifier when driven from a voltage source of 1000 Ω internal resistance.

(U.L.C.I., Electrical Technology 1B, May 1966)

8. Draw an equivalent circuit for a common-base transistor amplifier having a load resistance R_L. Derive expressions for the voltage and current gains in terms of the hybrid parameters and R_L, and hence show that the maximum power gain is obtained when

$$R_L = \sqrt{\frac{h_{ib}}{h_{ob}(h_{ib}h_{ob} - h_{rb}h_{fb})}}$$

(U.L. B.Sc. (Eng.) II, 1967)

9. Give sketches of the D.C. characteristics of a *p–n–p* junction transistor, and explain briefly the physical reasons for the shape of the curves.

The current amplification factor α for a grounded-base junction transistor at a frequency f is given approximately by

$$\alpha = \alpha_0\left(1+j\left(\frac{f}{f_\alpha}\right)\right)^{-1}$$

where α_0 is the low-frequency amplification and f_α is the cut-off frequency. Derive a corresponding expression for the current amplification factor in grounded-emitter configuration, and express the corresponding cut-off frequency in terms of f_α and α_0.

(U.L. B.Sc. (Eng.) II, 1966)

CHAPTER 10

COMMON TRANSISTOR CIRCUITRY

IN THIS chapter we describe, without great detail, some simple transistor circuits commonly met with in practice.

10.1. MULTI-STAGE AMPLIFIERS

In designing a voltage or current amplifier several points must be considered, including the attainment of the required gain (within a certain allowed distortion and a certain frequency range), the matching of input and output impedance, and the limiting of power dissipation. One usually knows the input signal available and the output required to feed, say, a power amplifier or relay, so one knows the gain required and the number of amplifying stages needed.

The separate amplifying stages have to be coupled together to make a multi-stage amplifier. (Remember that if the separate stage gains are A_1 and A_2, then the overall gain is A_1A_2, *not* A_1+A_2.)

The three basic interstage coupling methods are:

(a) capacitor–resistor coupling, commonly used in low-power audio-frequency amplifiers;
(b) transformer coupling, used for frequencies above about 100 kHz and for A.F. power stages;
(c) direct coupling, used in so-called D.C. amplifiers and other special circuits.

We will trace the design of a capacitor–resistor-coupled A.F. amplifier at this point, and return later to the other methods of coupling.

Suppose an A.F. amplifier is required to boost the output from a gramophone pick-up to a level adequate for driving a power stage, and a JJ1 transistor is to be used from a 15 V supply. Figure 10.1 shows a JJ1 transistor amplifying the pick-up signal and passing the output on to a second stage, whose input impedance is already known to be 1·5 kΩ. (We represent the second stage here by a resistor of $R_{\text{in}} = 1{\cdot}5$ kΩ.)

The D.C. load line

Let the D.C. load line pass through the $I_C = 3$ mA point (because this gives us a satisfactory operating point).

Then $$R_L+R_3 = \frac{15}{3\times 10^{-3}} = 5\ \text{k}\Omega \quad \text{(see Fig. 10.2)}$$

If we bias the base with a steady D.C. of 30 μA, the quiescent (no-signal) collector current will be 1·85 mA. Suppose we make $V_{BE} = 2$ V (which is $10\times V_{be}$) (see Section 9.8(b)),

then $$R_3 = \frac{2}{1{\cdot}85\times 10^{-3}} = 1080\ \Omega.$$

The nearest "preferred value" (value commercially obtainable) is 1000 Ω.

So $R_L = 5$ kΩ; $R_3 = 5$ kΩ-1 kΩ $= 4$ kΩ. The nearest "preferred value" is 3·9 kΩ. Then $R_L + R_3 = 1$ kΩ$+3·9$ kΩ $=$

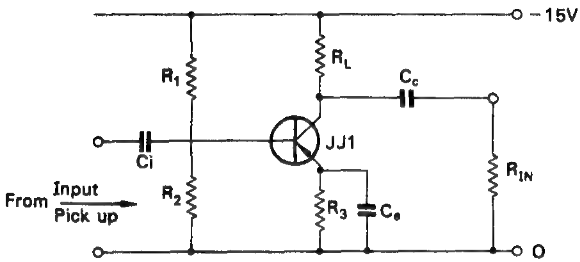


FIG. 10.1. An A.F. amplifier

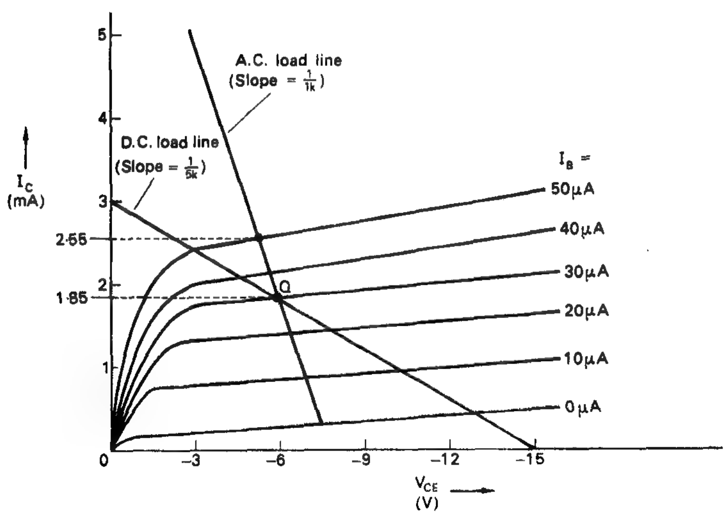


FIG. 10.2. Output characteristics of a JJ1 transistor (common emitter) with a D.C. and an A.C. load line

4·9 kΩ, and the I_C axis is crossed at

$$\frac{15}{4\cdot9\times10^3} = 3\cdot06 \text{ mA},$$

close enough to 3·00 mA for us to take the quiescent value of I_C as 1·85 mA.

The A.C. load line

The A.C. load line on the JJl transistor comprises R_L, R_{in} and $1/h'_{22}$ in parallel (assuming that the reactance of C_c is negligible). Also $R_L = 3\cdot9$ kΩ, $R_{in} = 1\cdot5$ kΩ, and we shall assume that

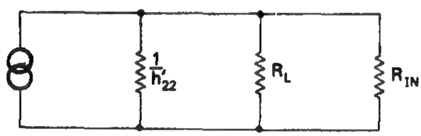


FIG. 10.3. Equivalent circuit for the A.C. load line

$h'_{22} = 50\times10^{-6}$ mho (although this value was actually determined from $I_B = 40$ μA, see Fig. 9.9).

We can represent the situation approximately by Figure 10.3, where

$$\frac{1}{h'_{22}} = \frac{1}{50\times10^{-6}} = 20 \text{ k}\Omega,$$

and can be ignored when in parallel with 3·9 kΩ and 1·5 kΩ.

So the load is $\dfrac{1\cdot5\times3\cdot9}{1\cdot5+3\cdot9} = 1\cdot080 \text{ k}\Omega = 1080\ \Omega.$

So the A.C. load line passes through the point Q and has a slope of 1/1080 mho. Hence an A.C. signal of 20 μA peak applied to the base will cause I_C to rise from 1·85 to 2·55 mA, i.e.

0·7 mA. But the current actually in the load is

$$\frac{R_L}{R_L+R_{in}}\times 0{\cdot}7\text{ mA} = \frac{3{\cdot}9}{3{\cdot}9+1{\cdot}5}\times 0{\cdot}7\text{ mA} = \frac{3{\cdot}9}{5{\cdot}4}\times 0{\cdot}7\text{ mA}$$

$$= 0{\cdot}5\text{ mA (peak).}$$

Bias stabilization

Choose R_2 such that $5R_3 < R_2 < 10R_3$. As $R_3 = 1\text{ k}\Omega$, let $R_2 = 6{\cdot}8\text{ k}\Omega$.

Then $$R_1 = \frac{V_{CC}-(I_CR_3+V_{BE})}{[(I_CR_3+V_{BE})/R_2]+I_B} \quad \text{(see Sect. 9.8(d))}$$

$$= \frac{15-(1{\cdot}8\times 1+0{\cdot}2)}{[(1{\cdot}85\times 1+0{\cdot}2)/6{\cdot}8\times 10^3]+30\times 10^{-6}}$$

$$= \frac{15-2{\cdot}05}{(2{\cdot}05/6{\cdot}8)\times 10^{-3}+30\times 10^{-6}}$$

$$= \frac{13}{294+30}\times 10^6$$

$$= \frac{13{,}000}{324}\text{ k}\Omega$$

$$= 40\text{ k}\Omega.$$

R_1 and R_2 in parallel give $\dfrac{6{\cdot}8k\times 40k}{6{\cdot}8k+40k} = \dfrac{272}{46{\cdot}8}k = 5{\cdot}8\text{ k}\Omega.$

The input resistance of a JJ1 transistor with a load of 1080 Ω is found from eqn. (9.14) to be about 3500 Ω. So to give a 20 μA peak swing in 3500 Ω requires $20\times 10^{-6}\times 3500 = 70$ mV.

But the bias resistors R_1 and R_2 take

$$\frac{70\text{ mV}}{5{\cdot}8\text{ k}\Omega} = 12\ \mu\text{A},$$

so the *total* input current (peak) $= 32\ \mu A$.

Hence *overall* gain $= \dfrac{0{\cdot}5\times10^{-3}}{32\times10^{-6}} = \dfrac{500}{32} = \mathbf{15{\cdot}7}$

The stability factor is

$$\frac{1k+5{\cdot}8k}{1k+(1-0{\cdot}98)\,5{\cdot}8k} = 6,$$

which is less than 10 and therefore satisfactory.

Capacitors

At low frequencies the reactance $1/wC_c$ of the coupling capacitor C_c increases considerably and so the gain of the multi-stage amplifier decreases as the frequency of a constant-voltage input signal is decreased (see Fig. 10.4).

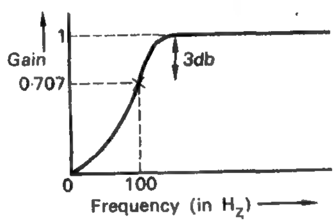


FIG. 10.4. The gain/frequency response curve of the JJ1 amplifier at L.F.

The frequency f at which the gain has fallen by 3 db* from its mid-frequency level is given by

$$\frac{1}{wC_c} = R_L + R_{\text{in}} \quad \text{where } w = 2\pi f$$

and the situation is represented in Figure 10.5.

* For decibel notation, see Appendix 4 in Vol. 1.

The value of C_c can be calculated if we assume that it is acceptable for the gain to be 3 db down at 100 Hz.

This gives $$\frac{1}{2\pi \times 100\, C_c} = 3{\cdot}9\ \text{k}\Omega + 1{\cdot}5\ \text{k}\Omega = 5{\cdot}4\ \text{k}\Omega$$

$$\therefore C_c = \frac{1}{2\pi \times 100 \times (5{\cdot}4) \times 10^3}$$

$$= \textbf{0·3 microfarad.}$$

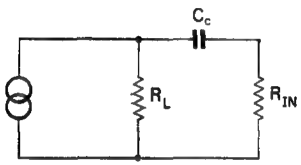


FIG. 10.5. Equivalent circuit for the L.F. cut-off

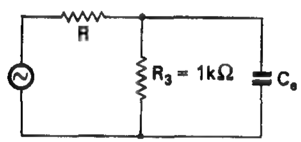


FIG. 10.6. Equivalent circuit for the emitter resistor R_3 and its by-pass capacitor C_e

The emitter resistor by-pass capacitor C_e simply provides an A.C. short-circuit across R_3, otherwise the bias would fluctuate with the signal. C_e has to "decouple" R_3 *and* whatever resistance is seen looking back into the transistor. Let us call this resistance R (see Fig. 10.6). The value of R is given by this formula:

$$R = \frac{\text{Input resistance} + \text{A.C. load}}{\beta + 1}$$

$$= \frac{3500 + 1080}{49 + 1}$$

$$= \frac{4580}{50} = 92\ \Omega.$$

The gain will be 3 db down when $1/wC_e = 92$. If the frequency is then 100 Hz, the value of $C_e = 1/2\times 100\times 92 = 17$ μF, and the nearest commonly available capacitor value is 16 μF.

The complete multi-stage amplifier is shown in Figure 10.7.

There is a decrease in gain at higher frequencies (compare the Miller effect in thermionics) due to (a) decrease in β because of physical effects, and (b) capacitive effects in the transistor and associated circuitry.

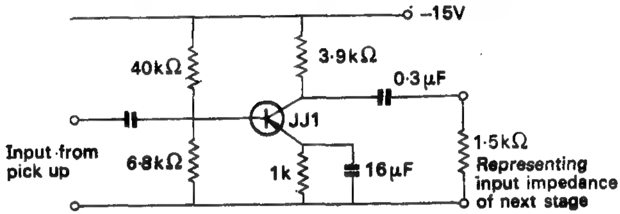


FIG. 10.7. The complete multi-stage A.F. amplifier designed around the JJ1 transistor

10.2. POWER OUTPUT STAGES

A signal from, say, a microphone, after being suitably amplified by a multi-stage amplifier (as in Section 10.1), is usually required to operate a loudspeaker, "modulate" a radio signal, work a recording disc cutter, or carry out some other task requiring considerable power. In order to be able to do so, the signal must first pass through a special amplifying stage called a "power amplifier".

The main way in which a power amplifier differs from the small-signal amplifiers previously considered is that now the signal amplitude extends over the greater part of the load line, and is no longer confined to a small part either side of the operating point.

The principles of power amplifier design are the same whether the stage involves transistors rated at 100 mW, or special ones able to dissipate 100 W. The common emitter configuration is nearly always used, as it gives best power gain.

Class A and class B types of operation are both popular. If both halves of the input signal remain on the linear region of operation of the transistor, the power stage is described as class A, but if only half of the input cycle falls in the linear region, it is called class B.

10.2.1. Single-ended power output stages

A typical transformer-coupled class A output stage is shown in Figure 10.8. Without the transformer the steady collector current would dissipate power continuously in the collector resistance. The operating point is limited by the peak collector

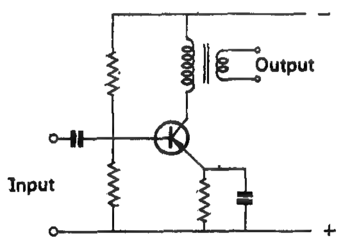


FIG. 10.8. A single-ended transformer-coupled class A output stage

voltage and current (set by the transistor itself), and the power dissipation (set mainly by the physical mounting of the transistor).

Consider a transistor whose output characteristics are sketched in Figure 10.9. The supply voltage is 10 V, and the maximum power dissipation allowed is 300 mW. (Dissipation greater

than this would cause damage to the transistor by the heat generated inside itself.)

In class "A" operation, the power taken from the battery is constant and shared between the load and the transistor, so, in the absence of a signal, when no A.C. power is developed in the load, the transistor suffers maximum power dissipation.

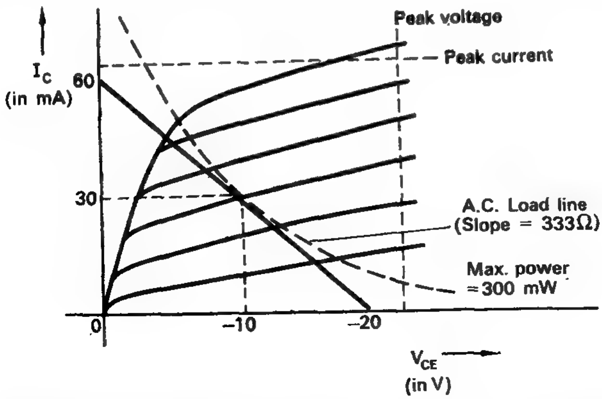


FIG. 10.9. Output characteristics of a transistor (common emitter)

If we neglect the resistance of the transformer primary, then the collector voltage = 10 V, and the collector current is 300 mW ÷ 10 V = 30 mA.

When a signal is applied, I_C must rise to 60 mA at zero collector voltage, and fall to zero when V_{CE} = 20 V. The load resistance clearly must be 20 V/60 mA = 333 Ω. If we know the resistance of the true load, in the secondary of the transformer, we can now find the required turns ratio in the usual way (Vol. 1, Sect. 9.7).

If $\beta = 50$ for the transistor used, and the output current is to swing ±30 mA, then the input drive current needed is approximately

$$\pm\frac{30}{50} = \pm 0{\cdot}6 \text{ mA}$$

The reader should note that this simple analysis neglects a few points, viz. V_{CE} has a finite "bottoming" value of about $\frac{1}{2}$ V*, which should be subtracted from the battery voltage; the input characteristic is non-linear particularly at low values of I_c, causing distortion; distortion is also caused by the reduction in the value of β at large values of I_c; and the biasing circuits cause a drain on the supply, causing a loss of p.d. which should also be allowed for (*V_{CE} when transistor is "full on").

10.2.2. Double-ended power output stage

The use of two transistors in class A push–pull has many advantages. Some of the distortion mentioned above is reduced. The calculations are similar to the foregoing. A typical circuit is shown in Figure 10.10.

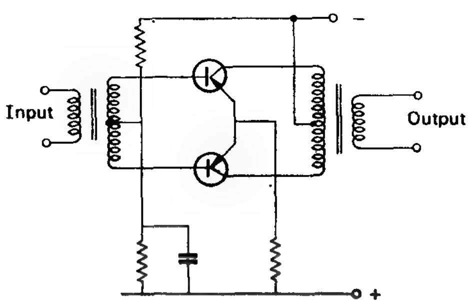


FIG. 10.10. Double-ended transformer-coupled class A push–pull output stage

In class B operation, two transistors must also be used, as each transistor handles only half of the output waveform. But the efficiency in practice (audio power out over battery power in) can be up to 70%, as compared to about 45% for class A. Also, when there is no signal, the battery drain current is negligible.

The most important requirement of such a circuit is that the two transistors must have equal current gains, so that both halves of the waveform are amplified equally. A typical circuit is shown in Figure 10.11.

But by far the most important development in this field, and one which has *no* counterpart in thermionic push–pull stages, is the use of a *pnp–npn* pair of transistors, known as a *complementary pair*, to make a transformerless push–pull output stage. (Any transformer introduces distortion and limits frequency range, so for a long time the idea of an output stage without one had been the unattainable dream of the audio designer.)

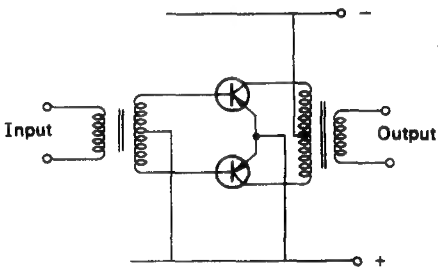


FIG. 10.11. Double-ended transformer-coupled class B push–pull output stage

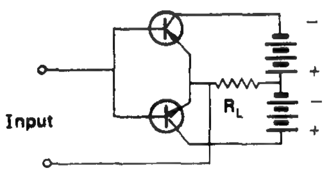


FIG. 10.12. Double-ended transformerless push–pull output stage using a complementary pair

The basic arrangement is shown in Figure 10.12. A positive going input signal is applied to both bases, causing the *n–p–n* transistor to conduct, and producing a positive output voltage

across the load R_L. The *p–n–p* remains non-conducting. But when the negative-going half of the input signal arrives, the *p–n–p* conducts and produces a negative output voltage across R_L. This time the *n–p–n* remains "OFF". Thus the application of an A.C. signal to the input produces an amplified signal across R_L, but in the absence of a signal no current flows in the load. As with a normal class B stage, a small bias must be applied to

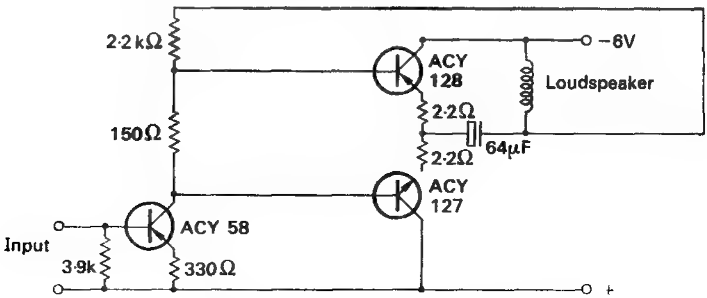


FIG. 10.13. Practical circuit of Figure 10.12

reduce "crossover" distortion (the distortion due to the fact that the transistor characteristics are not linear just at the switch-on points), and a practical circuit is shown in Figure 10.13. The ACY58 is called the "driver", and develops an alternating p.d. across the 150 Ω resistor. Half of this p.d. drives each half of the complementary pair consisting of ACY127 and ACY128. Each works as a common emitter amplifier, so the load (a small loudspeaker) receives pulses from each transistor in turn.

10.2.3. "Thermal runaway"; precautions

Whenever a transistor conducts, the dissipation of power (at the rate V_cI_c) at the base–collector junction causes an increase in the temperature of the junction. This in turn produces a rise

in the leakage current I_{CO} across the junction, which in turn adds to the rate of heat generation. This "vicious circle" can result in the destruction of the transistor by melting of the junction; or, if I_C is limited by an external resistance, it will cause an increase in distortion by allowing the transistor to "bottom" (reach saturation current) earlier in each cycle. To prevent this happening, I_C must be stabilized against changes in I_{CO}, and any heat generated at the junctions must be efficiently conducted away, to keep the working temperature low. For this purpose power transistors are often attached to "heat sinks" (see Appendix V).

10.3. TUNED AMPLIFIERS

Transformer coupling, based on a parallel resonant or "tuned" circuit, is commonly used above about 100 kHz and up to several hundreds of MHz. In contrast to the R.C. coupling of the multi-stage A.F. amplifiers of Section 10.1 (where the gain was kept as constant as possible over the audible range), with transformer-coupled tuned amplifiers the gain is constant only for a small band of frequencies.

The tuned circuit, inductance and capacitance in parallel, is in resonance at the mid-band frequency, and is connected in either the base or collector circuit. The inductance used is one-half of a transformer, the other half being in the adjoining amplifier stage. Thus coupling of one stage to the next is achieved, and so, at the same time, is impedance matching for maximum power gain.

At the resonant frequency the tuned circuit acts as a pure resistance, and, over the small frequency range in use, the transistor parameters are considered constant. High gain and good selectivity make tuned amplifiers very important in telecommunications.

10.3.1. Single tuned amplifier stage

Figure 10.14 shows a simple single-tuned amplifier stage. The input is applied to the base of the first transistor, which is loaded with a parallel-tuned circuit in the collector, so tuned as to be resonant at the frequency it is desired to amplify. It has a re-

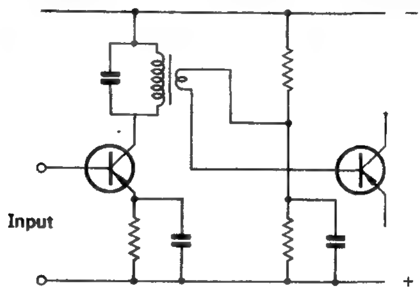


FIG. 10.14. Simple single-tuned amplifier stage

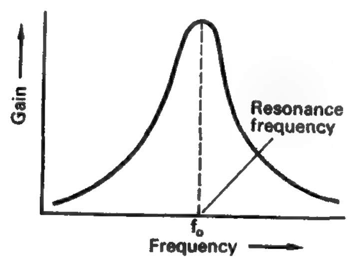


FIG. 10.15. Response curve of a single-tuned amplifier stage

sponse curve as shown in Figure 10.15. By means of the transformer the amplified signal is fed to the base of the next transistor, which is itself probably loaded with a tuned circuit, and so on. In practice a tapped inductance is commonly used, to allow feedback to be applied in opposition to the internal feedback which occurs naturally in transistors at high frequencies.

10.3.2. Double tuned amplifier stage

A more useful response curve can be achieved if the circuit of Figure 10.14 has a capacitor connected across the transformer *secondary*. We now have two tuned circuits, and the stage is called "double tuned". A typical response curve is shown in Figure 10.16, and a circuit in Figure 10.17. k is the coefficient of coupling between the two coils, and Q is called the quality

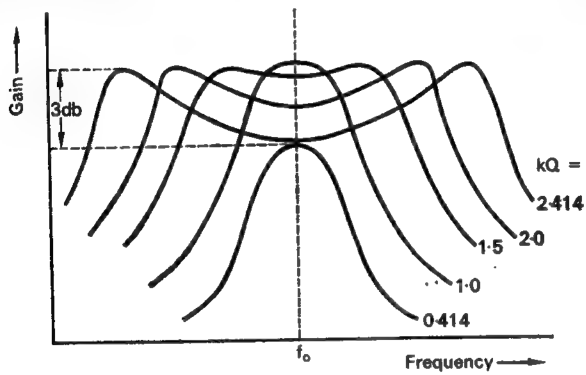


FIG. 10.16. Response curve of a double-tuned amplifier stage

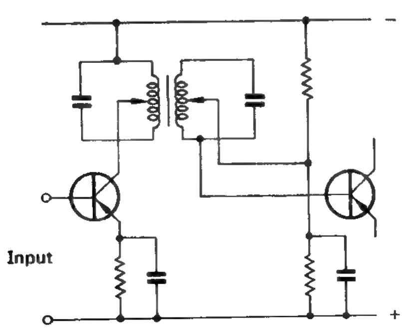


FIG. 10.17. Simple double-tuned amplifier stage

factor $\left(=\frac{\text{Reactance}}{\text{Resistance}}\right)$ of each coil; the two coils are usually identical. Figure 10.16 shows how choice of the quantity kQ can produce a response curve with steep sides and a flat top. The steep sides mean that frequencies near but not in the chosen band are considerably weaker; such an arrangement is desirable if a radio receiver is to handle all the audio frequencies transmitted by a certain station, but reject other nearby stations. Alternatively a very narrow bandwidth could be obtained, suitable for separating a morse transmission from another very nearby.

Circuits of the double-tuned type are to be found in every portable transistor receiver.

10.4. OSCILLATORS

The basic principles of oscillators were discussed in Vol. 1, Chap. 10. The reader will recall that if an amplifier circuit includes feedback and a resonant (tuned) circuit, oscillation will occur, and the D.C. powered circuit will give a sinusoidal, alternating output. However, if the conditions for oscillation are satisfied over a number or range of frequencies, a non-sinusoidal waveform will be produced. We will discuss briefly some transistor circuits of both classes.

10.4.1. Sinusoidal oscillators

(a) *The tuned collector feedback oscillator*

A common emitter arrangement is shown in Figure 10.18, and this basic circuit can be used over a wide range of frequencies, from low A.F. (using laminated-iron core transformers) up to

several hundred MHz (using radio-frequency coils). The feedback signal is obtained from the small secondary winding of the transformer, and, with 180° phase reversal (remember that a common emitter amplifier produces a 180° change), is applied back to the base of the transistor. By a suitable choice of turns ratio the overall gain can be made greater than unity; and *R* and

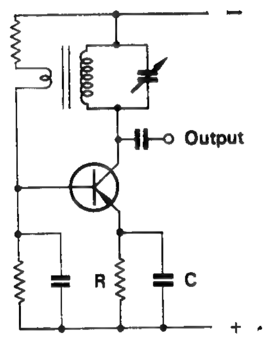


FIG. 10.18. Tuned collector feedback oscillator (common emitter)

C can be adjusted to limit the gain so as to ensure a good sine-wave output.

Similar principles can be applied to the common base configuration.

Hartley and Colpitts oscillators (see Vol. 1, Chap. 10) can similarly be built using transistors instead of valves as the active elements.

(b) Resistance-capacitance feedback oscillators ("phase shift" oscillators)

A resonant circuit is not essential in a sinusoidal oscillator circuit; the tuned circuit can be replaced by a suitable R.C. network in the feedback path, as shown in Figure 10.19, a very

useful single frequency sine-wave oscillator. The R.C. circuit produces a phase shift which depends on frequency, and at a certain frequency this shift is 180°. So oscillations occur at this frequency, provided the transistor stage gain is greater than or equal to the attenuation produced by the R.C. network. This can be shown to be approximately 29, so a transistor with a β of

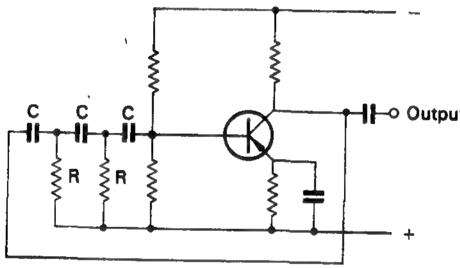


FIG. 10.19. Phase shift oscillator

at least 50 is usually needed. The frequency of oscillation is given by

$$f \doteqdot \frac{1}{\sqrt{6} \times 2\pi CR}$$

10.4.2. Relaxation oscillators

There are many types of relaxation oscillator, and all produce waveforms which are *not* sinusoidal. They include multi-vibrators, saw-tooth generators, blocking oscillators, and others. Here we shall deal only with the multi-vibrator.

A typical circuit is shown in Figure 10.20, and consists of a two-stage amplifier with symmetrical cross-coupling. At any instant, one transistor is bottomed and the other "off"; a moment later the situation is reversed. The output, taken from either collector, is a rectangular waveform which can easily be

made square, or symmetrical. The action is as follows: the transistors T_1, T_2 cannot have *identical* values of β so, when the circuit is first switched on, the current rises faster in one transistor than in the other. Suppose this occurs in T_1. The voltage at the collector of T_1 falls, and C_1 conducts this decrease to the base of T_2, causing the current through T_2 to decrease. Therefore the voltage at the collector of T_2 rises, and C_2 couples this rise

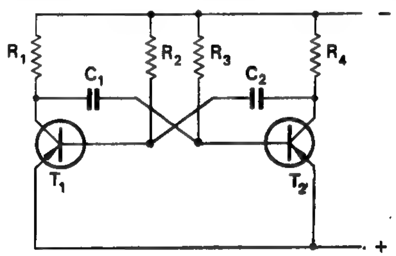


FIG. 10.20. The multi-vibrator relaxation oscillator

to the base of T_1, increasing the current through T_1 even further. This process continues until T_1 is fully switched on (bottomed) and T_2 is fully switched off. C_1 is now able to start charging through R_3 until the base of T_2 has gone far enough negative to start T_2 conducting again. Thus T_2 turns on and T_1 turns off; thereafter C_2 gradually charges up through R_2 and the whole cycle of events repeats itself.

The voltage waveforms of this astable multi-vibrator are shown in Figure 10.21. ("Astable" means that it has no stable condition. Slight circuit changes can produce a monostable or a bistable multi-vibrator.)

A useful "rule of thumb" for the time of one part of the two-part cycle is P_1 (sec) $= C_1R_3$ (megohms$\times$microfarads). Likewise, the time for the other part of the cycle is approximately $P_2 = R_2C_2$.

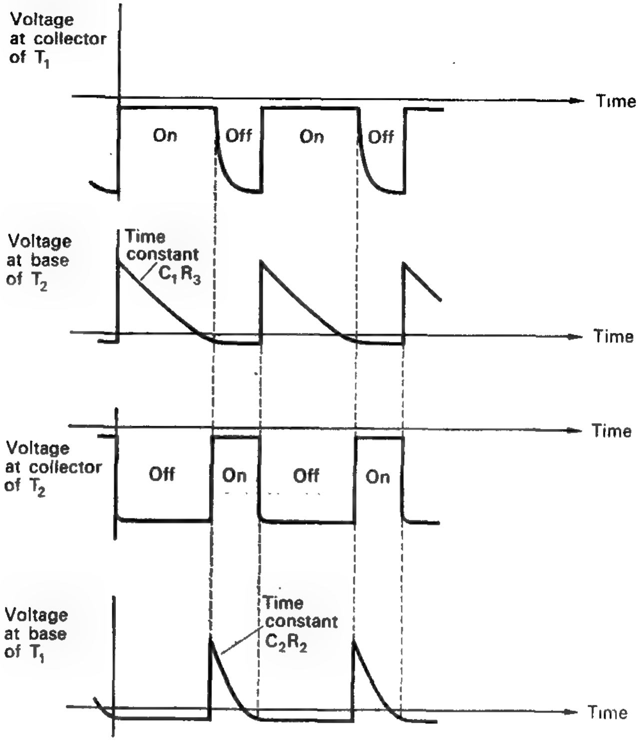


Fig. 10.21. Voltage waveforms of the astable multi-vibrator shown in *Figure 10.20*

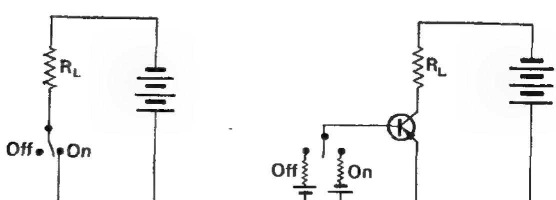


Fig. 10.22. Transistor as a switch

10.5. THE TRANSISTOR SWITCH

In Section 10.4.2 we saw how two transistors are connected as a multi-vibrator: in this application each transistor is behaving literally as a switch; it is either "on" or "off". Figure 10.22 shows this analogy. When switched to the OFF position the base–emitter junction is reverse biased (as well as the base–

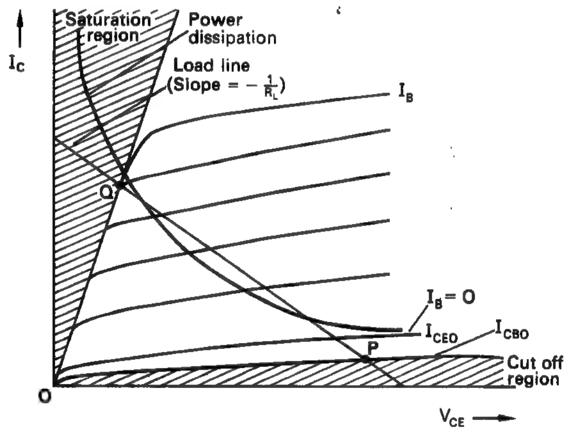


FIG. 10.23. The output characteristics load line and power dissipation curve for the transistor operating as a switch

collector junction) and only the very small collector–base leakage current I_{CBO} flows through the load R_L. This current can be as low as 1 nA with silicon planar transistors, giving an "off" resistance of about 1000 MΩ. Even with germanium alloy transistors it can be as high as 1 MΩ. In this condition the transistor is said to be "cut off" (see point P in Fig. 10.23). When switched to the ON condition the base–emitter junction is forward biased and the transistor conducts. The resistance associated with it is only

a few ohms, so the voltage drop across it is very small. At point Q in Figure 10.23, further increase in I_B would *not* cause any increase in I_C, and the transistor is said to be "bottomed" or "saturated". In this state both junctions are forward biased.

The usual method of operation uses a rectangular waveform applied at the base, to switch the transistor very quickly. As the operating point moves from P to Q it must momentarily pass through the region of high-power dissipation, hence the necessity for speed. Typically a transistor can switch some 10^6 times faster than a domestic light switch, and a large collector current can be switched by a small base current.

The reader should note that if the transistor is to be used to switch an electromagnetic relay (or other inductive load), the induced back e.m.f. could damage the transistor, and a diode (sometimes with capacitor) has to be connected across the coil to afford a safe path for the resultant current.

10.5.1. Logic circuits

Electronic computers make extensive use of basically simple transistor switching circuits. Two basic decisions are needed in computer logic: AND NOT. Similary OR or NOT can be used. Sometimes NOR gates alone are used. A simple AND gate is shown in Figure 10.24: if the first, second, and third input signals are all present together, a single output results. If one or more inputs are not present, there is no output.

An OR gate gives a single output if one or more of the three inputs are present. In both these circuits the polarities of inputs and output are the same, i.e. the inputs must be positive-going, and the output, if any, will be positive-going.

The output polarity can be reversed simply with a common emitter stage, called a NOT circuit. A combination of AND and NOT circuits is called a NAND circuit, and a combination

of OR and NOT a NOR circuit. A simple NOR circuit is shown in Figure 10.25. The reader will notice that NAND and NOR gates are in fact simpler than AND and OR gates.

Also extensively used in computers are bistable multi-vibrators, used in binary counting circuits. Figure 10.26 shows a simple

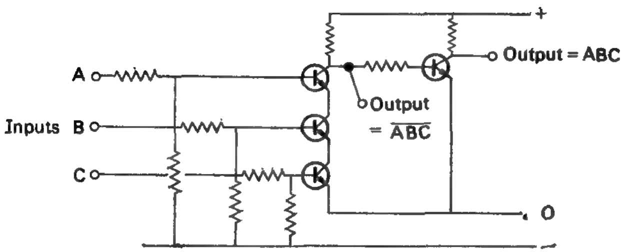


FIG. 10.24. An AND gate: NAND followed by NOT gives AND

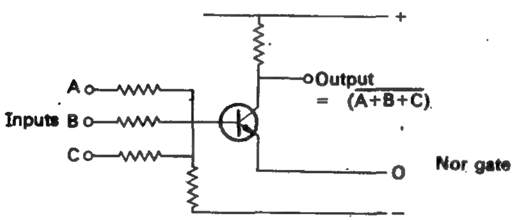


FIG. 10.25. Basic NOR gate

example; the circuit will stay with either transistor on and the other off, but the arrival of an input pulse will reverse the situation. So two input pulses produce one output pulse, and the effect of division by two is evident. (This type of "divide by two" circuit also finds extensive use in electronic organs, because the frequency of, say, *C* below middle *C*, is exactly half that of middle *C*.)

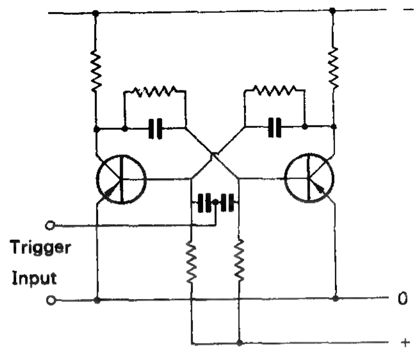


FIG. 10.26. Simple bistable multi-vibrator

10.6. D.C. AMPLIFIERS

The principles of D.C. amplifiers, and the design problems associated with them, were discussed in Vol. 1, Chap. 10. Here we will merely indicate how transistors can be used similarly, with advantage, instead of valves.

10.6.1. Directly-coupled circuits

In Figure 10.27 the resistor R_4 allows the base of T_2 to operate at the same voltage as the collector of T_1. But because of the loss of signal introduced by the negative feedback across R_4, a better arrangement is to replace R_4 by a Zener diode with suitable bias (i.e. to connect a resistor from the + supply line to the emitter of T_2).

An ingenious circuit, called the Darlington pair, has a very high input impedance and high gain, making it very suitable for low noise input circuits. It is in effect a direct-coupled double

emitter follower, as shown in Figure 10.28. R_1 adjusts the bias by controlling the base current to T_1. Some semiconductor manufacturers offer a composite unit containing effectively the two transistors in one capsule.

Transistors can also be used in the basic compensating or "balancing" circuit sometimes called a "long-tailed pair".

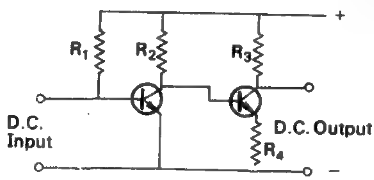


FIG. 10.27. A two-stage directly-coupled amplifier

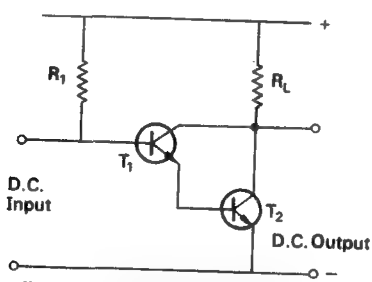


FIG. 10.28. A directly-coupled double emitter follower amplifier (a "Darlington pair")

A single-ended version is shown in Figure 10.29. The problem of obtaining two identical transistors T_1, T_2 can be overcome by forming two transistors in the same wafer of silicon, or by taking two wafers from the same diffused slice and mounting them side by side in the same encapsulation. This ensures that changes due to temperature change and supply fluctuation are fully cancelled out.

The development of the field effect transistor has proved most important to D.C. amplifier techniques, because of the inherently

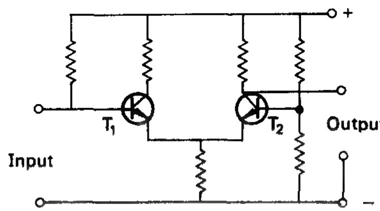


FIG. 10.29. A "long-tailed pair" D.C. amplifier

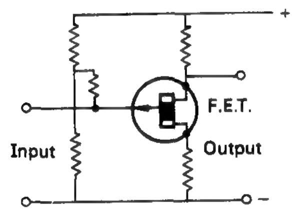


FIG. 10.30. A field effect transistor D.C. amplifier

high input impedance associated with them. Figure 10.30 shows the first stage of a high input impedance amplifier with low noise performance. Typically the input impedance is 10 MΩ, the capacitance only a few picofarads, and the noise figure about 1 db when fed from a high impedance source. Typical frequency response might be zero (i.e. D.C.) to 10 kHz. The output from such a stage could go direct to the input of the long-tailed pair of Figure 10.29.

10.6.2. Chopper circuits

An alternative approach to D.C. amplification is to "chop" the D.C. signal into a square wave A.C. signal, amplify this by conventional circuits, and then rectify the output. This can be advantageous because an A.C. amplifier is not subject to cumulative drift.

The disadvantages of mechanical methods of chopping are overcome by using a transistor as a switch. The principle is shown in Figure 10.31. The D.C. signal acts as the collector supply voltage; the base is fed with a square wave A.C. The transistor is thereby switched from cut-off to saturation and back at the square-wave frequency. When switched off, the collector

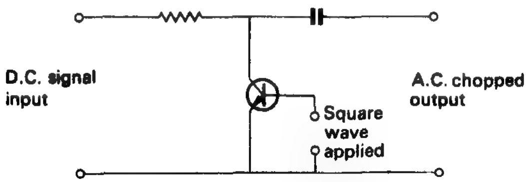


FIG. 10.31. A "chopper circuit" D.C. amplifier

voltage equals that of the D.C. input; when "on", the transistor "shunts" the D.C. signal. The output signal, therefore, is an alternating one, of frequency equal to that of the square-wave drive, and of amplitude nearly equal to that of the D.C. input.

10.7. INVERTERS

Just as A.C. can be changed to D.C. by means of a rectifier (see Sect. 6.2), so D.C. can be converted to A.C. by what is called an inverter circuit.

A typical inverter circuit is shown in Figure 10.32. The two transistors T_1 and T_2 are power types, and the transformer is of the saturable reactor type, the magnetic material of the core having a near-square hysteresis loop. The polarities of the windings must be as shown, and the reader should satisfy himself that the circuit will oscillate. The output will be square-wave. Remember that at the moment of switching on, one transistor will con-

duct more than the other, because they cannot be entirely identical. Having produced A.C., and, if necessary, stabilized its frequency, it is easy to "amplify" it if more power is required. The efficiency of conversion can reach 90%. Having thus obtained A.C., it can be brought to the desired voltage by another transformer, and then, if desired, rectified by fairly conventional circuits. If 50 Hz A.C. is not specifically required, it is

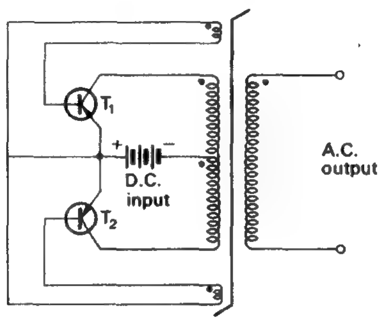


FIG. 10.32. A basic inverter circuit

worth remembering that higher inverter frequencies will allow the use of physically smaller transformers and smoothing components.

Very low-power inverters are to be found in the electronic flash units used by photographers; typically, a 9 V hearing-aid battery supplies the power to charge a capacitor to about 500 V to work the flash tube. Larger inverters (15 W) are used to work electric shavers from car batteries, and models giving more than 1 kW are available for such purposes as emergency reserve supplies.

EXPERIMENT 10.1: The transistor as a switch.

Apparatus required:

12 V 2 A D.C. supply.
12 V 24 W lamp and holder.
0–1 A meter.
Resistor, 100 Ω, 5 W.
Transistor type OC26 or OC35, on heat sink.

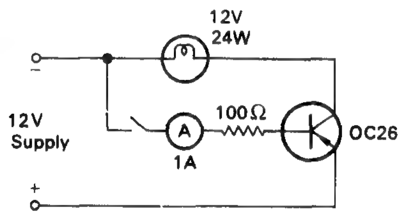


FIG. 10.33.

Procedure. Connect up the circuit of Figure 10.33. Note that the lamp is switched on and off by means of the base connection, although not more than 200 mA flows in this part of the circuit. Thus, a large direct current could be controlled with a small switch.

EXPERIMENT 10.2: An audio-frequency oscillator (common emitter).

Apparatus needed:

Transistor type OC81.
Output transformer intended for two OC81s in push–pull to match 3 Ω.
Resistor, 4·7 kΩ.
Resistor, 10 Ω.
Loudspeaker, 3 Ω.
Supply, 5 V D.C.
Oscilloscope.

Procedure. Set up the circuit of Figure 10.34. A powerful note around 1000 Hz should be heard. Connect the oscilloscope input earth to the supply +, and examine the collector, base, and emitter waveforms by connecting the input lead to each in turn.

Further projects. Connect the oscilloscope across the loudspeaker. Replace the 10-Ω resistor with a variable one of up to 50 Ω, and look at and listen

to the effect of varying it. A lower resistance gives less "negative feedback", more output, and more distortion. Distortion is heard as an increasing "richness" to the note, i.e. more of the higher harmonics. It is seen as further departures from the ideal "sine-wave" shape. Too high a resistance stops the oscillations.

Try also the effect of capacitors of between 0·001 and 0·1 F connected across the primary of the transformer. They will lower the fundamental note.

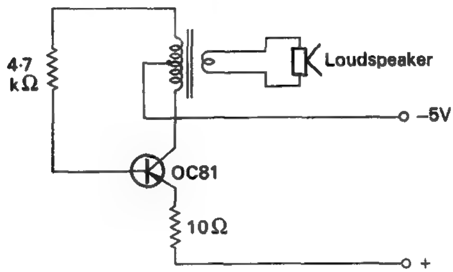


FIG. 10.34.

EXPERIMENT 10.3: A radio-frequency oscillator (common base)

Apparatus needed:

Transistor type OC45.
Long-wave tuning coil with reaction winding.
Variable tuning capacitor (say 0–350 pF).
Resistor, 4·7 kΩ.
Resistor, 5 Ω.
2 capacitors, 0·01 F.
Supply, 5 V D.C.
Portable radio with long waveband.

Procedure. Set up the circuit of Figure 10.35. Switch on the receiver and stand it within a foot of the circuit, and set the tuning to a spot on the long-wave band where no station is received. Adjust the variable capacitor; at one or two points the receiver should indicate by an obvious change of "background noise" that the oscillator is functioning at the selected frequency. If nothing is heard, reverse the connections to the reaction winding (if either coil is the wrong way round, the oscillator will not function). (For those using "Radionic" kit, which is ideal for many of these experiments, the long-wave coil colours have been marked on the diagram.)

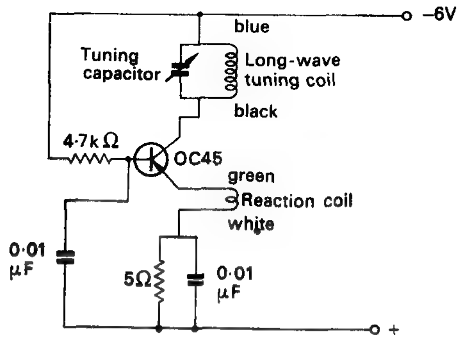


FIG. 10.35.

Note. It would be illegal to attempt to make the signal so generated travel any distance, e.g. by connecting any form of aerial.

The "amplifying" part of this circuit has the input to the emitter, and the output from the collector. The base is earthed, as far as A.C. is concerned, by the capacitor connecting it to earth. This is an example of a "common base" circuit, which finds more use at radio frequencies.

QUESTIONS

1. Figure 10.36 shows two methods by which a transistor amplifier can be biased so that its operating point is $V_{ce} = 3$ V and $I_c = 3$ mA. (Assume $V_{be} = 0{\cdot}7$ V, each transistor has $\beta = -50$, and the cut-off current I_{cbo} is negligible.) Explain, with numerical examples where relevant, the relative merits of the two amplifiers in practice.

(U.L. B.Sc. (Eng.) Part 1, 1967, Part question)

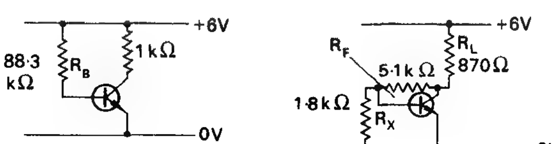


FIG. 10.36.

2. Describe briefly an audio-frequency class A transistor output stage. Indicate the necessary operating conditions, and give approximate values of collector efficiency that may be obtained.

3. A simple tuned-collector transistor amplifier is shown diagrammatically in Figure 10.37. The transistor common-emitter output characteristics are given by:

	$V_{ce} = 0$	-2	-4	-6	-8	-10	Volts
$I_b = 0$	$I_c = 0$	-1	-1	-1	-1	-1	mA
$I_b = -50$ A	$I_c = 0$	$-3{\cdot}5$	-4	$-4{\cdot}5$	-5	$-5{\cdot}5$	mA
$I_b = -100$ A	$I_c = 0$	$-6{\cdot}0$	-7	$-8{\cdot}0$	-9	-10	mA
$I_b = -150$ A	$I_c = 0$	$-8{\cdot}5$	-10	$-11{\cdot}5$	-13	$-14{\cdot}5$	mA

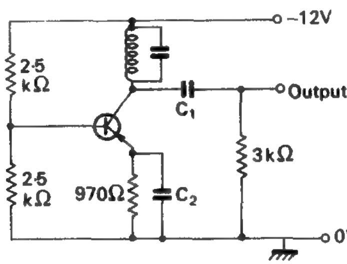


FIG. 10.37.

The tuned circuit has a dynamic resistance of 940 Ω, the D.C. resistance of the coil is negligible, and the capacitors C_1 and C_2 have negligible reactance at signal frequencies. Assuming that the magnitude of the base-emitter voltage is negligible, find the no-signal operating point of the transistor.

4. Figure 10.38 shows the circuit of a transistor Colpitts oscillator which, when suitably designed, will produce nearly sinusoidal oscillations. Derive the condition for marginal oscillation, and the frequency of this oscillation, in terms of the circuit elements and the parameters of the transistor. State clearly any assumptions you make.

If the condition for oscillation is satisfied by a sufficient margin, oscillations will eventually be limited in amplitude. What mechanism can be responsible for such limitation in the circuit of Figure 10.38? What is the basic requirement of any method of amplitude limitation?

(U.L. B.Sc. (Eng.) Part 2 (Electrical) 1967)

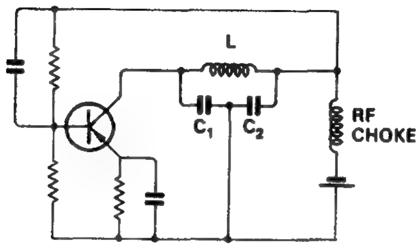


FIG. 10.38.

5. Write an essay on the transistor direct-coupled amplifier, giving particular attention to the method of characterizing drift, the main sources of error, and the means by which these errors can be reduced.

(U.L. B.Sc. (Eng.) Part 2 (Electrical) 1967)

6. A certain application requires an astable multi-vibrator, with an output waveform having a peak-to-peak amplitude of approximately 10 V, a periodic time of 10 sec and a mark/space ratio of 1/2. Using approximations, appropriate to a first trial design, devise a suitable circuit using two identical transistors, assuming that appropriate transistors are available. Explain the function of each circuit component. State any assumptions made concerning the properties of the transistors. Indicate which transistor will stay in the ON condition for the longest time (i.e. 1 sec per cycle of operation).

(U.L. B.Sc. (Eng.) Part 2 (Electrical) 1967

7. Explain briefly how the circuit shown in Figure 10.39 could operate as a binary unit. Show how the operation of the unit could be improved by the addition of diodes. Determine the smallest value of R_L for which the circuit will function if the current gain h_{fe} for the transistors is 30.

(I.E.E., Part 3, June 1967

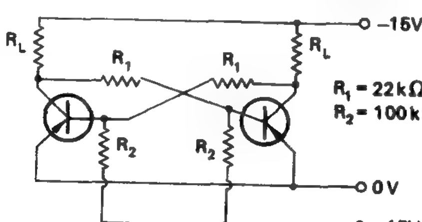


FIG. 10.39.

8. State briefly the problems associated with the direct-coupled amplifier. Explain with the aid of a circuit diagram how complementary transistors have aided the design of such amplifiers. Show that the circuit of Figure 10.40 may be considered as a single transistor. Derive expressions for the short-circuit current gain and the input resistance. Evaluate these expressions if the short-circuit current gain for each transistor $h_{fe} = 60$ and $h_{ie} = 1\ \text{k}\Omega$.

(I.E.E., Part 3, June 1967)

9. Compare the various classes (A, B, etc.) of amplifier and explain why class B is commonly used in the output stage of audio-frequency push–pull power amplifiers. Show that the theoretical maximum efficiency of a typical

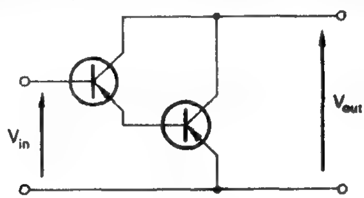


FIG. 10.40.

class B push–pull stage is about 78%. Show also that the transistor dissipation increases by about 50% when the input signal level is reduced by 36% from the value for maximum efficiency. Comment on this phenomenon.

(I.E.E., Part 3, Dec. 1967)

10. Indicate the manner in which circuits for performing the AND and OR functions may be built up using resistor–transistor logic. Use a truth table to show the relationship between the inputs and outputs of an equivalence logic circuit and indicate how this circuit may be built up from the resistor–transistor logic circuits already described.

(C. and G.: Advanced Telecomms and Electronics, Digital Computers, May 1967, adapted)

11. Draw the circuit of a collector-coupled bistable multi-vibrator, using two similar transistors, suitable for use in a multi-stage binary divider to divide 31,250 Hz to 50 Hz. Give typical component values and include the interstage coupling components in the circuit diagram. Draw a block diagram showing the interconnection of the minimum number of stages required to achieve this division ratio.

(C. and G., Supp. Telecomms and Electronics, Television Broadcasting I, June 1967)

CHAPTER 11

SPECIAL DEVICES, PROCESSES AND USES

11.1. HETEROJUNCTIONS

We noted in Section 6.1.3 that it is possible to make heterogeneous crystals of intermetallic-compound semiconductors, i.e. the junction is formed between two *n*-type regions (say) of different electrical conductivities, and different energy band structures. If the distance through the crystal over which the change takes place is very small, then a p.d. will exist across this heterojunction just as in the conventional *p–n* junction, because the forbidden energy gap on each side of the junction is different. In Figure 11.1, representing an *n*-type heterojunction, equilibrium is reached when the Fermi levels of the materials on each side of the junction line up. Under reverse-bias conditions the energy or potential "spike" is greater and few electrons have sufficient energy to surmount it and move from right to left; but under forward bias the height of the spike is reduced and many electrons can flow from left to right. As a rectifier the heterojunction is not as good as a *p–n* junction because the ratio of forward current to reverse current is not great; however, the current flow is exclusively electrons (in the *n*-type heterojunction considered here) and recombination with holes, a rela-

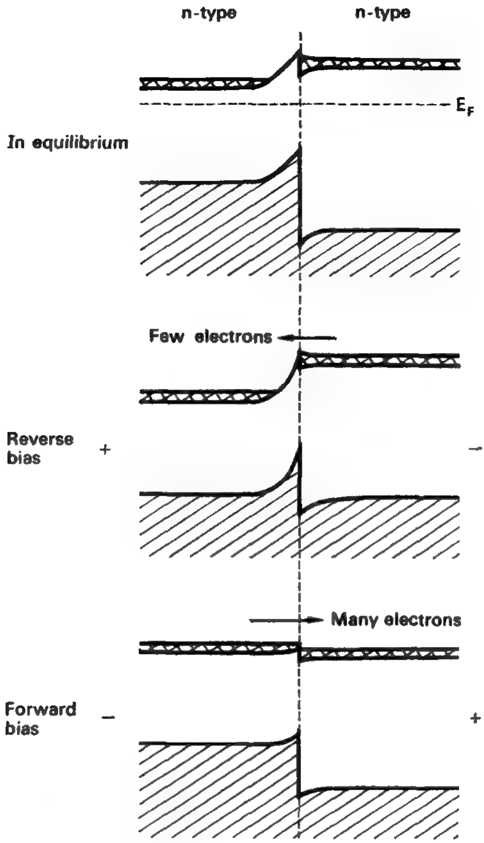


FIG. 11.1. An *n–n* heterojunction

tively slow process, is not necessary as it is with a *p–n* junction. Consequently the heterojunction finds use in very high-speed rectifying applications such as are to be found in digital computers.

It is possible to combine a *p*-type wide energy gap material and an *n*-type narrow energy gap material as a *p–n* heterojunction, which has important optical properties (see Section 11.6.7).

11.2. SONAR AMPLIFIER

Ultrasonic waves in some compound semiconductors, such as cadmium sulphide, can be amplified by a novel process. This is because electrons flowing through a semiconductor, especially at high current densities, *do* cause a very small motion of the atoms of a crystal. But a sound wave flowing through the crystal does this as well. If a voltage is applied across a semiconductor crystal of sufficient value so that the velocity of electrons flowing is equal to the velocity of sound in the crystal, then the sound wave can experience amplification. Such a "sonar amplifier" can produce a very high amplification of an ultrasonic wave directly. Because of the high current densities and consequent heating effect, it is usual to pulse the voltage, giving the crystal enough time between pulses to cool down.

11.3. RADIATION DETECTION

If the width of the depletion layer in a germanium *p–n* junction is considerably increased from a typical value of 10^{-4} m or less, up to about 7 mm (by a technique called "lithium drifting"), the diode makes a very useful β-ray detector in nuclear radiation research. Typically a volume of about 50 cm^3 can separate β-ray energies up to 5 MeV to within 1%, a resolution better than that obtainable with conventional gas ionization detectors or scintillation counters.

A thin depletion layer silicon *p–n* junction in reverse bias is capable of α and β radiation detection. There is, however, a very real chance of irreversible damage to the junction by the radiation bombardment.

11.4. POWER CONTROL

A very simple way of controlling D.C. power is shown in Figure 11.2. This is most inefficient, the surplus energy being dissipated as heat in the resistance. Figure 11.3 depicts the characteristics associated with this basic method of D.C. control. (See also Expt. 11.1.)

A semiconductor-operated switch in series with the motor is an alternative method of control. Figure 11.4 represents this, from which we see that it is the ON–OFF ratio of the switch that controls the motor speed. Figure 11.5 depicts the characteristics of this method. The motor has a large inductance associated with it, and the square wave of Figure 11.4 is effectively smoothed. Because of this inductance a large voltage pulse can be

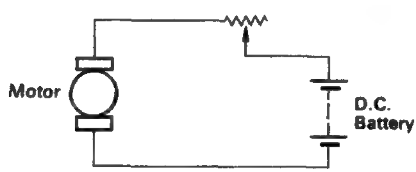


FIG. 11.2. Basic D.C. power control circuit

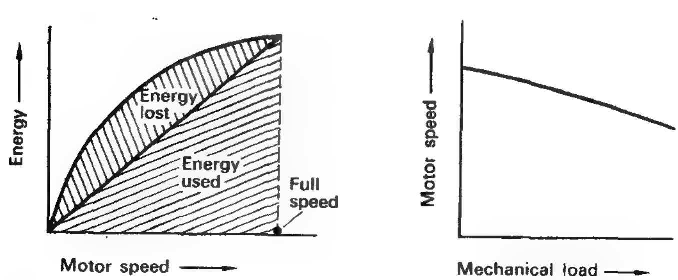


FIG. 11.3. Characteristics associated with Figure 11.2

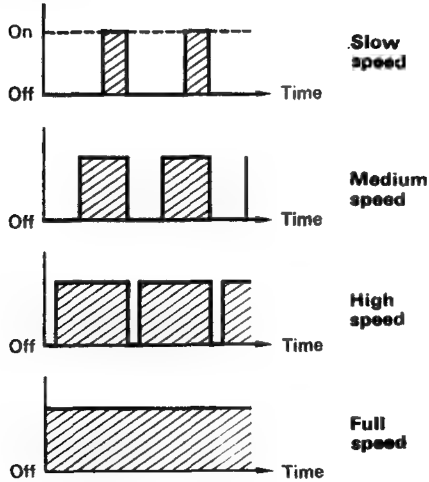


FIG. 11.4. Characteristics of power control by switching

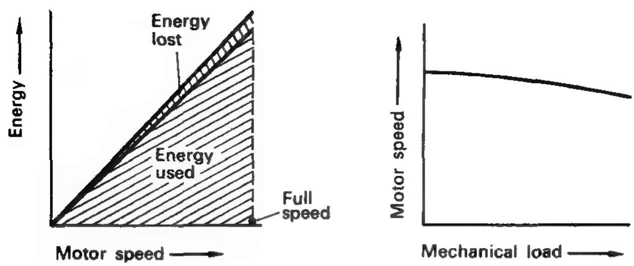


FIG. 11.5. Characteristics associated with Figure 11.4

generated when the current switches off, so it is necessary to include a rectifier across the motor winding to limit the induced voltages to the value of the supply voltage, so that the semiconductor control circuit is not damaged.

The semiconductor control circuit can either use power transistors or *thyristors:* the latter are preferable at high powers, whereas the former need simpler control circuits.

11.4.1. The thyristor

The thyristor, or silicon-controlled rectifier (SCR), is a fast-acting solid state switch, analogous to the gas-filled triode or thyratron. The thyristor is available with current ratings of up to about 300 A. It has proved to be of great importance industrially, in controlling A.C. power. The basic thyristor characteristic is shown in Figure 11.6. Normally it blocks current flow in both directions, but when suitably triggered by a pulse it allows forward current (the characteristic then looking like that of a conventional silicon rectifier) whilst still blocking reverse current. As soon as forward current stops for any other reason, it returns to the "off" state, and requires to be triggered again. The thyristor is a four-layer device (e.g. *pnpn*), and is

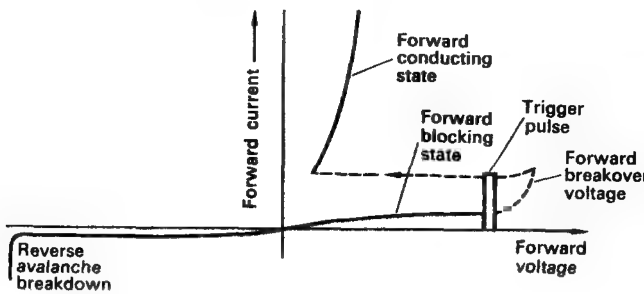


FIG. 11.6. Basic thyristor characteristic

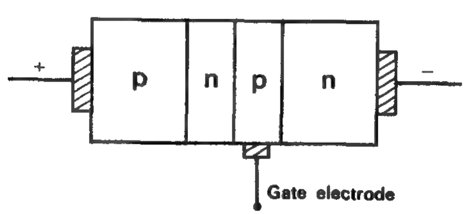


FIG. 11.7. A thyristor or four-layer device

shown in Figure 11.7. The pulse of current that "flips" the thyristor into the forward conducting state is applied to the *gate* electrode. It flips back into the blocking state when the forward (load) current falls to almost zero.

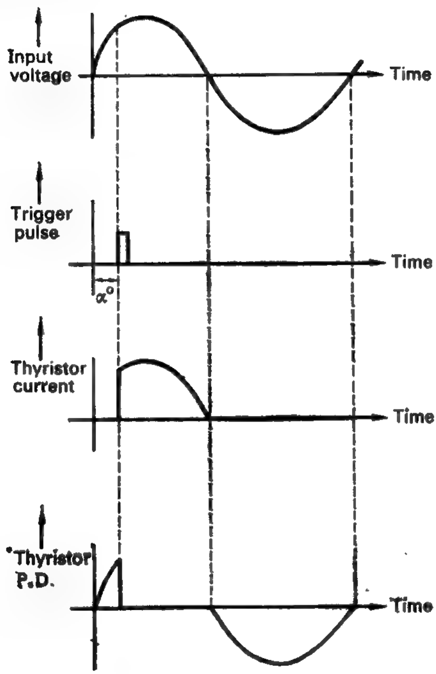


FIG. 11.8. Characteristics of a thyristor in use

Thus the thyristor can control the output current of a rectifier circuit, if the timing of the "trigger" pulse in the operating cycle is varied (see Fig. 11.8). The angle α is called the firing angle. A typical trigger pulse is about 3 V at 100 mA for about 10 μsec. Note from Figure 11.8 that *removal* of the trigger pulse has no effect on the thyristor output—it can *only* by turned off by

removal of the main current. This is analogous to the situation in a thyratron (Vol. 1) in which the grid voltage has no control over the load current once the valve is conducting: the only way to switch it off is to decrease the anode voltage. In a thyratron the "switching-off" time is dependent on the deionization time for the gas filling; this is limited, whereas a thyristor has a switching-off time of 10–100 μsec depending on the current.

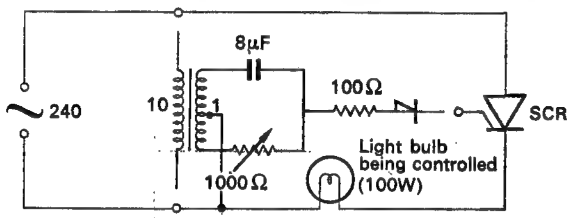


FIG. 11.9. A thyristor as a simple half-wave rectifier controlling power

A single thyristor operates over only half a cycle, as seen in Figure 11.8. To function on both cycles an inverse parallel pair is necessary.

A simple half-wave controlled rectifier circuit is shown in Figure 11.9.

The 8 μF–1000 Ω circuit introduces a phase change so that the triggering pulse lags behind the "main" current. The amount of lag is controlled by the 1000-Ω variable and the main current does not flow until the triggering pulse arrives. When the main current falls below the "breakdown volts" (Fig. 11.6) it is cut off until triggered again. Thus the variable resistance controls the current through the lamp.

The great advantage of the thyristor is that the power is only supplied as needed, none being wasted in control resistances as in Figure 11.2. Thyristors are commercially available for from

3 to 300 A, and cost much less and last much longer than mercury arc rectifiers and ignitrons.

Other multi-layer devices now exist, of which the *triac* is one. It has five layers (*npnpn*) and is similar to two thyristors back-to-back. One triac can provide full-wave control, allowing current to flow in both directions of the A.C. cycle, whereas two thyristors would be needed (see Expt. 11.2).

Not all multi-layer devices are triggered from a third electrode. A four-layer diode is a two-terminal version of the thyristor with a small area, relying on a self-breakover for switching it into a conducting state. Typically it might switch into conduction at 50 V and 100 mA. It is used in switching circuits.

11.5. MICROWAVE DEVICES

Microwave devices are amplifiers and oscillators capable of working in the range 2 to 200 GHz (2000 to 200,000 MHz).

Normal transistors when used at very high frequencies suffer from the inevitable complications of capacitance and skin effect. Even specially designed transistors will only work up to about 4 GHz. To exceed this in the past the usual procedure has been to generate harmonics of this frequency by loading the oscillator with a non-linear reactance such as a reverse-biased *p–n* junction. Such a junction acts as a voltage-variable capacitor, and is the "Varactor" diode of Section 6.5.

A *single* device capable of sustaining microwave oscillations must simulate "negative resistance". The *Gunn effect* in gallium arsenide is a bulk property of the material, in which the electrons move across a strip of it in bunches called domains. The arrival of a domain at the collector is indicated by a sharp rise in current. If these pulses occur at a high enough frequency we

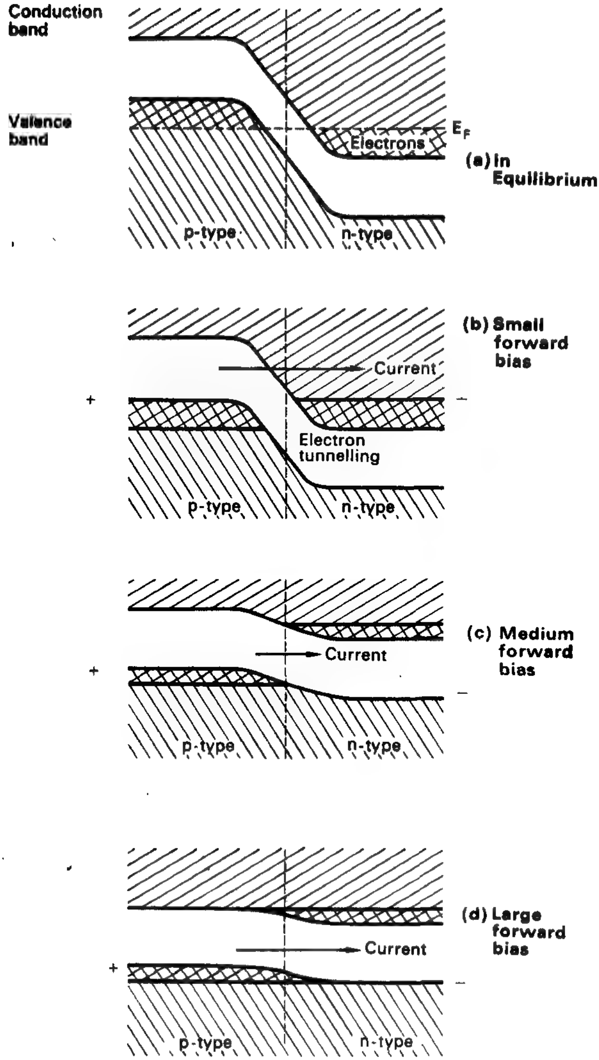


FIG. 11.10. Energy band diagrams for a tunnel diode

can produce microwave oscillations; the frequency is dependent on the thickness of the specimen. Typically, a piece 10 μm thick is capable of producing oscillations of 10 GHz, and such a thickness can be produced by epitaxial deposition.

The *tunnel diode* is a *p–n* junction that exhibits a negative resistance over part of its *V–I* current–voltage characteristic, and if it is biased in this region it can be made to oscillate at up to about 100 GHz. It is made by using heavily doped *p*- and *n*-type materials. When the impurity concentration is high, the impurity atoms become quite close together, and their associated energy levels merge and become indistinguishable from the conduction and valency band levels. So the concentration of carriers in the energy bands are high, and the associated Fermi level lies within the valency band for the *p*-type, and within the conduction band for the *n*-type. Figure 11.10(a) shows the energy band diagram for such a *p–n* junction in equilibrium. The width of the junction is very small (typically 3–30 nm) because of the high impurity concentration, and it is possible, under a small forward bias, for electrons in the conduction band of the *n*-type to undergo a quantum mechanical "tunnelling" process, and re-emerge in the valency band of the *p*-type material. As the forward bias is increased more levels in the *p*-type valency band are available to electrons in the *n*-type conduction band, and the current increases with forward bias voltage up to a certain maximum (Fig. 11.10(b)). If the forward bias is increased beyond this point the current decreases, because the electrons in the *n*-type conduction band are now raised above the holes in the *p*-type valency band, and less "tunnelling" is possible. But a limit is reached when the electrons and holes are completely out of line (Fig. 11.10(c)), and then any increase in forward bias voltage beyond this results in an *increase* in current because of forward injection, just as in the normal *p–n* junction (Fig. 11.10(d)).

The current–voltage characteristic of a tunnel diode is shown

in Figure 11.11, where the negative resistance region can clearly be seen. If the tunnel diode is biased in this region then it can compensate for the I^2R energy loss in the resistance of a tuned circuit which is otherwise capable of oscillating freely. Because tunnelling is a very fast atomic phenomenon, tunnel diodes can work at frequencies of up to 100 GHz.

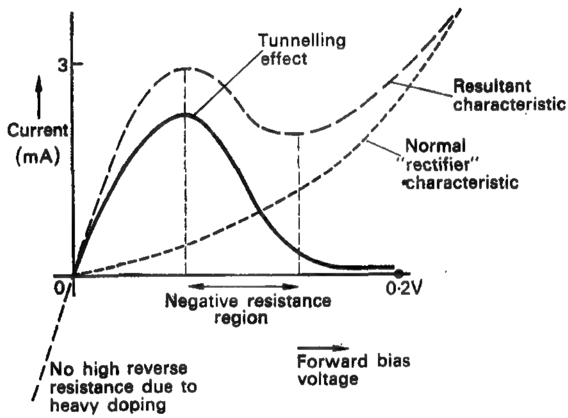


FIG. 11.11. The characteristic of a tunnel diode showing the negative resistance region

The power requirement for the tunnel diode in Figure 11.11 is about 1 mW. They can be germanium or silicon alloy junctions, or made of gallium arsenide or indium antimonide. In suitable circuits they can be used as microwave amplifiers or fast switches. They were discovered by the Japanese physicist Esaki in 1958.

The parametric amplifier, using a parametric diode or varactor, is also very useful at microwaves, and was mentioned in Section 6.5.

11.6. OPTO-ELECTRONICS

In conventional electronics, electric signals are conveyed along wires. In opto-electronics, signals are sent along light beams. Light beams can cross each other without confusion, and the information in them can be stored on photographic material. A transmitter is required capable of rapidly varying the intensity of the emitted light, and a receiver must incorporate detection and amplification.

11.6.1. Light emitting diodes

The "semiconductor lamp" is just a *p–n* junction operated with forward bias. Minority carriers are injected and recombine with some of the majority carriers, and release a certain amount of energy. In silicon and germanium the energy released is usually in the form of heat, but in gallium arsenide or phosphide it is in the form of light. This carrier recombination is a fast process, so the light-emitting diode does not have the frequency limitations associated with the heating of a wire filament or the ionizing of a gas, both of which are relatively slow.

The materials used for the *p–n* junction govern the amount and colour (wavelength) of the emitted light. Intermetallic compounds such as GaAs and GaP are popular, and can be doped with other elements to produce other colours. Typically a GaAs wafer 1 mm square might be forward biased at 1 V and 100 mA, emitting 50 μW of infra-red light (because of diffused zinc) at room temperature. If the junction is cooled in liquid air, the light power output is increased by a factor of about 10.

11.6.2. The laser

It should be clear that each electron transition from the conduction to the valence band (in materials suitable for light emission as in the last paragraph) is random and independent of all others; consequently the photons of light energy are emitted at random. Such a light source is described as "incoherent", as are hot filament bulbs and fluorescent tubes.

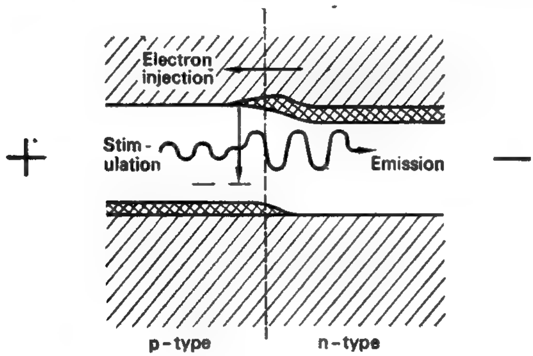


FIG. 11.12. Energy band diagram of a laser

However, if light of the emission wavelength is concentrated at the *p–n* junction, then the time of photon emission can be considerably influenced, so that the photons are now emitted at such times as to add constructively ("in step") to the original incident beam, so making it more intense, or amplifying it. In these circumstances the *p–n* junction is acting as a *l*ight *a*mplifier, and it is doing so by *s*timulating the *e*mission of *r*adiation. Hence the name "laser". Figure 11.12 shows this on an energy band diagram for a forward-biased *p–n* junction.

We can now dispense with the original, separate, incident light beam that produces the stimulation, and make the laser

action "self-starting". If a *p–n* junction has highly polished and perfectly parallel sides perpendicular to the plane of the junction, then a photon emitted from a single transition, travelling parallel to the junction, will be reflected back and forth repeatedly through the junction region, stimulating *other* transitions. These photons will be similarly aligned, and the intensity of the light will build up, whilst remaining "coherent". Note that those photons emitted *not* parallel to the junction receive no such amplification, so their effect quickly decreases to negligible proportions. The resulting coherent light beam eventually passes out of one side of the crystal, because one side is purposely made an imperfect reflector.

The coherent light from a laser is of exactly one particular wavelength only, and is strictly parallel. The latter property means that such light may be converged by a convex lens to a spot the diameter of which is of the order of one wavelength, and consequently the energy per unit area in the spot will be extremely high—high enough to melt the most refractory materials with great precision. The uses of this property extend from machining to surgery.

By modulating the light output of the laser (by applying an electric signal to vary the junction current), we have a communications system whose handling capacity is enormous. 10^9 television signals (bandwidth 5 MHz) could be carried on one laser beam! The implications for communications of the future are obvious.

11.6.3. Optical transistor

It is possible to couple together a light-emitting diode (Sect. 11.6.1) and a photocell (Sect. 6.7) into a four-terminal device in which the input is electrically isolated from the output. This is called a semiconductor relay, and a typical arrangement is

shown in Figure 11.13. The transmitter is a forward-biased p–n junction (of low resistance) and the receiver a reverse-biased p–n junction (of high resistance.) The signals are transmitted very rapidly by the light beam. Because a light beam is used, rather than the relatively slow drift of charged carriers, the high-frequency performance is good. The gain comes from the fact that it is an impedance converting device. The transmitter and

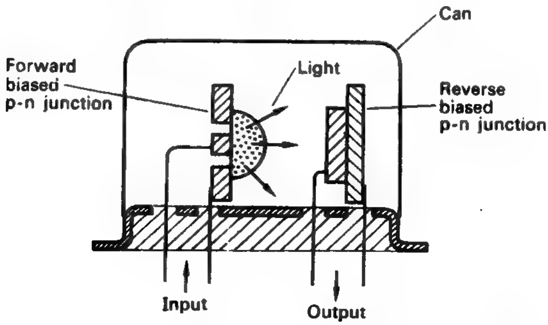


FIG. 11.13. Optical transistor or semiconductor relay

receiver can be separated by quite large distances, the light beam being carried by an optical fibre. A bundle of optical fibres in an "insulating" (light-tight) cover is sold under the trade-name of "CROFON", and looks exactly like a length of insulated single flex. (See Expt. 11.3.)

11.6.4. Electroluminescence

This is the general name given to the direct conversion of electrical to optical energy in a solid; the action of the light-emitting diode (Sect. 11.6.1) is just one form.

A semiconductor powder such as zinc sulphide is mixed in a thin plastic film and placed between two flat electrodes, one of which is transparent. The application of a voltage can inject

carriers into the powder grains, each of which behaves like a *p–n* junction. Recombinations occur and light is emitted. The process is not as efficient as in a single *p–n* junction, because each powder grain does not contain a perfect junction. However, the luminous intensity is even over the whole area of the material, making the device useful in instrument panels, night lights, and displays.

The efficiency is greatest for green light and electroluminescent panels usually emit this colour. Other single colours can be produced by choosing the right impurity atoms, so that the loss of energy of the carriers in recombination corresponds to the wavelength of the desired colour. White light is produced with a mixture of red, green, and blue producing powders.

11.6.5. The phototransistor

This is a junction transistor in which the can and contents are transparent, so that light can fall on the region near the collector. As the base–collector junction is reverse biased, incident radiation increases the collector current, the transistor amplifying the effect.

An important point is that phototransistors (and photodiodes) have very small active areas, consequently they are very useful as position indicators. (See Expt. 11.4.)

11.6.6. Photo-avalanche diode

The photo-avalanche diode has almost superseded the phototransistor. In this diode, the incident light energy creates hole–electron pairs which, on accelerating, create others, and so on. The consequent current can be quite large, although triggered by an extremely weak light pulse. The thermionic equivalent is the photomultiplier (Vol. 1), capable of a gain of up to 10^6 by the secondary emission of electrons.

11.6.7. Heterojunctions

It was mentioned in Section 11.1 that some heterojunctions have important optical properties. Consider a *p*-type material of wide forbidden energy gap, forming a junction with a narrow-gap *n*-type material. The incident electromagnetic radiation, if visible light, has a shorter wavelength than that associated with the wide forbidden energy gap, and consequently passes through the *p*-type material unattenuated. But in the *n*-type material it is absorbed, because its wavelength is now similar to that associated with the narrow forbidden energy gap, and so hole–electron pairs are created in the close vicinity of the junction. They are removed by the electric field before they can recombine. This results in a very efficient solar cell.

11.7. THE UNIJUNCTION

The unijunction transistor (or double base diode) is a three-terminal device like an ordinary transistor, but it has only *one* *p–n* junction. Its construction and symbol are shown in Figure 11.14.

Metallic connections are made at the ends *B* and *C* of an *n*-type silicon bar whose resistance is about 5000 Ω. A voltage

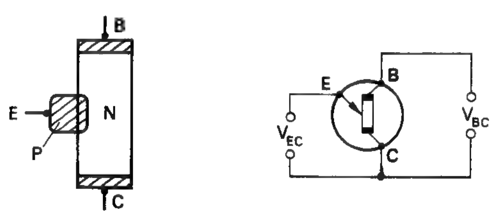


FIG. 11.14. The unijunction transistor (or double base diode).

V_{BC} is maintained across the bar. A p-type emitter E is diffused into the side of the bar between B and C, and this forms a rectifying junction. If E is open-circuit, a fraction of V_{BC} appears across E to C ($V_{BC'E}$). By applying a second external p.d. between E and C (V_{EC}), the emitter is reverse biased if $V_{BC'E}$ is greater than V_{EC}, and only a small leakage current flows. The emitter becomes forward biased when V_{EC} is greater than $V_{BC'E}$, and a hole current flows into the bar, increasing its conductivity and reducing $V_{BC'E}$. So as I_E increases, V_{EC} decreases, giving the negative resistance characteristics shown in Figure 11.15.

This negative resistance property can be used in a very economical relaxation oscillator circuit, shown in Figure 11.16. A saw-tooth waveform is generated across C and a pulse wave-

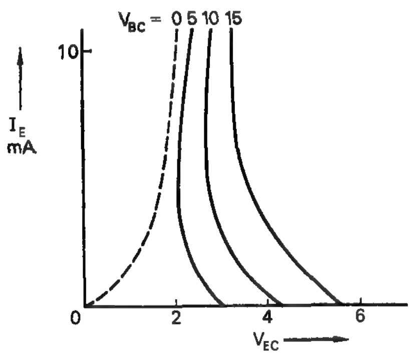


FIG. 11.15. Characteristics of a typical unijunction

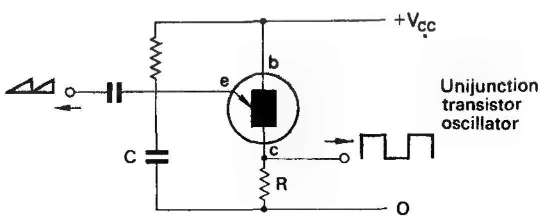


FIG. 11.16. A unijunction relaxation oscillator

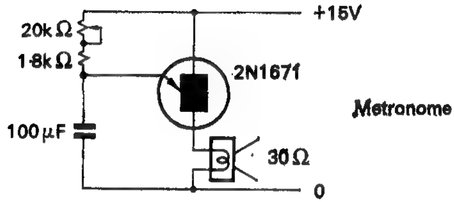


FIG. 11.17. Circuit for an acoustic metronome using a unijunction. (Courtesy of R. G. Hibberd)

form across *R*. The circuit for an acoustic metronome is given in Figure 11.17, variation of the 20 kΩ resistance changing the frequency of the "clicks" given out by the loudspeaker. (See Expt. 11.5.)

11.8. FIELD EFFECT TRANSISTORS

The planar diffusion technique, briefly described in Section 8.1.6, enables a different type of transistor to be made. The "field effect" transistor, or FET, which dates from 1928, is very important in its own right as a single electronic component; but, as we shall see in Section 11.9, it is revolutionizing electronics in the rapidly growing field of microminiaturization. By "field effect" is meant the change in carrier concentration in a semiconductor due to variation of an applied field.

11.8.1. The junction FET

This is basically like the unijunction of Section 11.7, but it has *two p–n* junctions across the bar. Its construction is shown in Figure 11.18. The junction connections are called "gates", and the ends of the bar "source" and "drain". A bar of *p*-type semi-

conductor material is shown, with two *n*-type regions formed on opposite sides of the bar about midway. If these two *p–n* junctions are reverse biased, then their depletion regions extend into the bar (by an amount dependent on the value of reverse bias) leaving a "channel" through which the conduction from

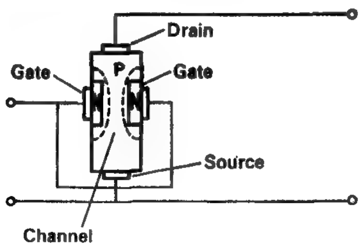


FIG. 11.18. The junction field effect transistor

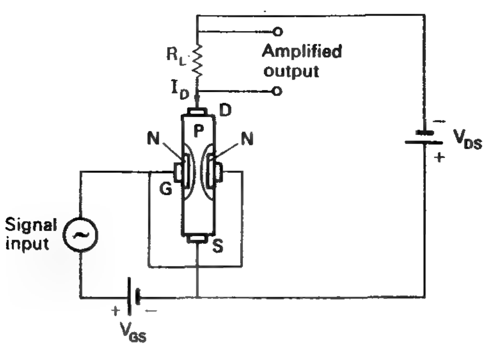


FIG. 11.19. The junction FET as an amplifier

source to drain must flow. Clearly then, the effective resistance of the bar can be controlled by the reverse bias on the gate. The channel produced is wedge-shaped because there is an external p.d., applied between drain and source (V_{DS}), and so there must be a potential gradient in the bar between drain and source.

Figure 11.19 shows a junction FET as an amplifier. The exter-

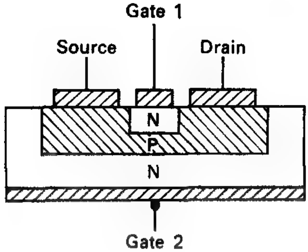


FIG. 11.20. The planar diffusion junction FET

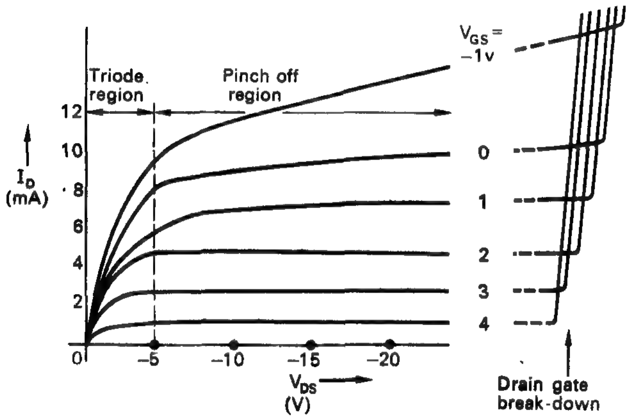


FIG. 11.21. The characteristics of a typical planar diffusion junction FET

nal p.d. of V_{DS} establishes a current through the load resistance R_L, and the *p*-type bar between drain *D* and source *S*. Another external p.d., V_{GS}, biases the *n*-type gates, and a certain channel width exists in the *p*-type bar, allowing a certain steady drain current I_D to flow through R_L. If now a signal is applied, the gate bias and the channel width vary likewise, and so does I_D; consequently the voltage across R_L will vary with the input signal, but with many times the magnitude. As the input signal is applied to a *reverse*-biased *p–n* junction, the input impedance of the

junction FET amplifier is very high, e.g. for a silicon planar junction FET it can be 20 MΩ or more.

Figure 11.20 shows a cross-section of a junction FET made by the planar diffusion process, and Figure 11.21 shows the I_D/V_{DS} characteristics for various values of V_{GS}. On the characteristics shown, when V_{DS} exceeds about 5 V there is little increase in current; this is known as "pinch-off at the drain". It must be emphasized that the explanations given here are necessarily greatly simplified. (See Expt. 11.6.)

11.8.2. The insulated gate FET, or metal–oxide–semiconductor transistor ("MOST")

The junction FET described in Section 11.8.1 is a "depletion" device—the electric charge density in the channel can only be decreased by increasing the reverse bias on the gate. In the "insulated gate" FET, which dates from 1962, the charge density in the chan-

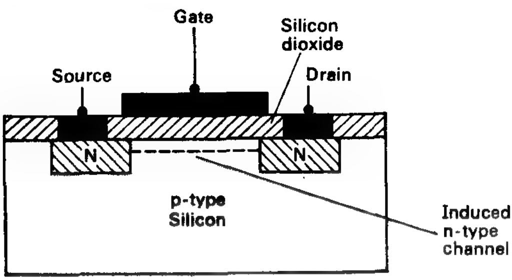


FIG. 11.22. An *n*-channel insulated gate FET or metal–oxide–semiconductor transistor (MOST)

nel can be increased by inducing electrostatically a conducting channel between the two electrodes, this process being called "enhancement". The MOST is available in four types: *p*-channel depletion or enhancement, and *n*-channel ditto. Figure 11.22 shows the construction of an *n*-channel MOST, the chan-

nel being formed in the surface of the *p*-type silicon slice by electrostatic induction from the gate potential across the silicon dioxide insulating layer, into the silicon. The gate electrode is usually aluminium, hence the name "*m*etal–*o*xide–*s*emiconductor *t*ransistor".

The drain is kept positive, and the source electrode is often shorted to the substrate (to ensure that the *n*–*p* junctions do not become forward-biased). If the gate electrode is negative, positive charge is induced in the channel, the electrons being repelled

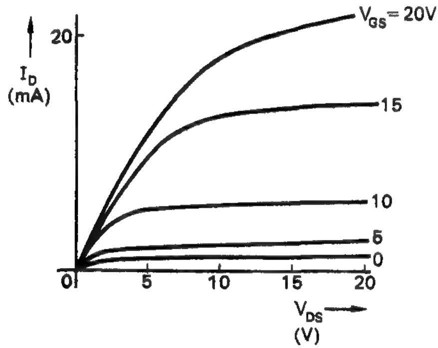


Fig. 11.23. Characteristics of a typical MOST

from this region. Thus the channel width is reduced; this is the depletion mode of operation, and—like the junction FET of Section 11.8.1—there is a shallow diffused junction in the surface of the silicon. If, however, during construction of the MOST, *no* junction is formed, then, in the absence of a potential on the gate, the source and drain are separated by a very high resistance region and the source-drain current is very small (provided that V_{DS} is less than the breakdown voltage of the region). But if a positive potential is applied to the gate, the *p*-type surface of the silicon will be converted to *n*-type by electrostatic induction across the silicon dioxide insulating layer. Majority carriers in

the silicon are forced away from the surface, and minority carriers attracted to it from the bulk of the semiconductor material. This is the "enhancement" mode of operation, and it is important to note that in this case no depletion layer is associated with it. Figure 11.23 shows the I_D/V_{DS} characteristics for various values of V_{GS} for a typical MOST.

It should be clear that, as the gate electrode is separated from the channel in the semiconductor by a layer of insulator, the input impedance to the gate is *extremely* high, of the order of $10^{12}\ \Omega$.

The FET functions by majority carrier action in the semiconductor bar alone, in direct contrast to the conventional bipolar transistor which uses minority carriers in the base region and majority carriers in the emitter and collector regions.

The FET has distinct advantages over the conventional transistor for most applications: it, is, however, more expensive.

A most important fact is that although FETs can be used individually, their outstanding advantage comes in "integrated" circuits (see next paragraph). For instance, switching circuits, flip-flops, etc., can be made up entirely of MOSTs, with no other component such as capacitors and resistances at all.

11.9. MICROELECTRONICS

Throughout the history of electronics there has been a marked need to make the individual components of a circuit as small as possible, and to increase the density of packing by such techniques as printed circuits, plug-in potted modules, etc. The culmination of this trend is microelectronics, which incorporates thin and thick film techniques, monolithic silicon integrated circuits (SICs), multiple chip circuits, and hybrid types, all under

the general name of "integrated circuits". Whereas in the past the electronic components were connected together to form an electronic circuit, in microelectronics it is scarcely possible to take part of a circuit and say "this is one component". Here we can give only the barest outline of the approach and possibilities.

Consider the simple amplifier circuit of Figure 11.24. This can be thought of in semiconductor terms as Figure 11.25(a), where the three resistors are pieces of suitably doped *p*-type

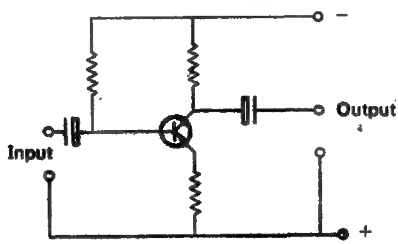


FIG. 11.24. A simple amplifier circuit

silicon, the two capacitances are the self-capacitance of a *p–n* junction, and these components are connected by wires. The arrangement of Figure 11.25(b) follows when it is realized that two pieces of the same type of semiconductor need not be connected by a wire; and Figure 11.25(c) follows by subtle geometric juggling. Here the whole circuit is combined in one single semiconductor crystal, containing the right *P*, *N*, and *I* (insulating) regions. Such an integrated circuit would be manufactured by the usual silicon planar diffusion process; more complicated multi-layer devices are made by the epitaxial process.

Alternatively the passive components (resistances, capacitors) can be produced by thin and thick film deposition techniques, which have recently been developed to produce active devices such as the FET.

In order to microminiaturize an electronic circuit the designer will normally first redraw it without inductors (as these cannot yet be satisfactorily miniaturized), and then perhaps adopt a hybrid approach, using thick and thin film techniques for the resistors and capacitors, and the planar process for the transistors.

The advantages claimed for integrated circuits include lower cost (for very large production runs), increased reliability and performance, and smaller size (although the latter is often a minor advantage).

A whole range of microelectronic devices are now commercially available, to undertake a vast range of electronic tasks.

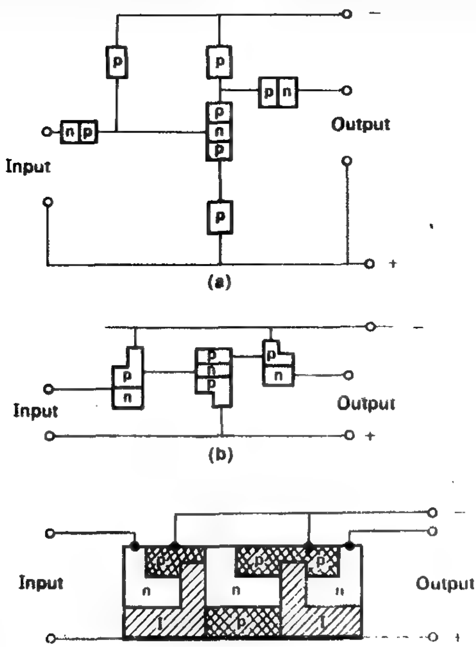


FIG. 11.25. Stages in the miniaturization of the circuit in Figure 11.24

EXPERIMENT 11.1: Transistor control of D.C. motor.

Set up the circuit of Figure 11.26 in which the motor is of the type supplied with "Malvern" energy-conversion sets, i.e. suitable for driving models, and requiring up to 10 V D.C.

The friction of the motor itself is quite sufficient "load" to allow the speed control principle to be demonstrated.

The advantages to note are (1) that the variable resistance by which speed is controlled is a small, radio-type such as would burn out if used *directly*

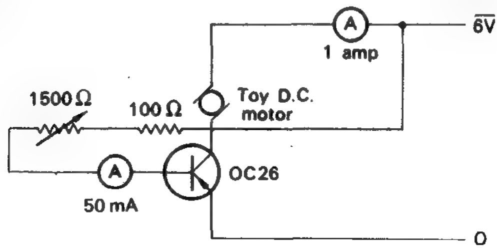


FIG. 11.26.

to control the motor; (2) that stalling the motor by overloading it (stop the shaft with the fingers) does not cause a dangerous increase of current through the variable resistance. It does, of course, increase the heat dissipated by the transistor.

EXPERIMENT 11.2: Triac (double thyristor) control of A.C.

Apparatus required:

Triac type 40432 (RCA; Proops).
Potentiometer, 250 kΩ 1 W.
Resistor, 10 kΩ 1 W.
2 capacitors, 0·1 F, 450 V d.c.
Lamp, 240 V 1000 W (e.g. theatrical spot or floodlamp, or projector lamp with usual cooling arrangements).

Procedure. Wire the lamp, triac, and control circuitry to the A.C. mains as in Figure 11.27. Every precaution must be taken to prevent anyone touching any part of the circuit. Adjustment of the 250 kΩ potentiometer will control lamp brightness from full on to off. A voltmeter connected across the lamp will indicate *roughly* the average voltage applied to the lamp.

To study the waveform of the voltage applied to the lamp, either the lamp circuit or the oscilloscope must be isolated from the mains. This cannot easily be done for the lamp circuit, as an isolating transformer capable of handling 1000 W would be needed; but the oscilloscope can

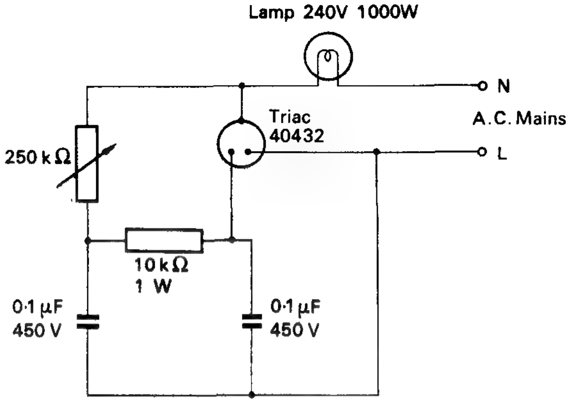


FIG. 11.27.

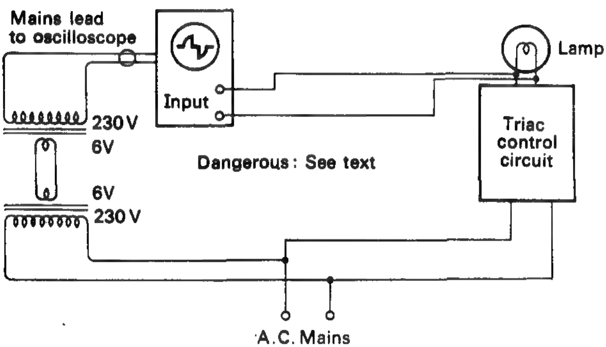


FIG. 11.28.

easily be isolated by using back-to-back two low-voltage transformers of the kind found plentifully in school and other laboratories. Set up the circuit of Figure 11.28, *but take great care, as any part of the oscilloscope could now be "live"*. Study the way the lamp voltage waveform varies when the brightness is adjusted using the potentiometer.

The non-sinusoidal lamp current waveform is necessarily rich in harmonics of the 50 Hz fundamental frequency. These harmonics extend well into the range of radio frequencies, consequently considerable "interference" to radio reception is both radiated, and carried along the lamp and mains

leads. To show this, set the lamp to about half brightness and bring a switchedon portable receiver near to the circuit. Radiated interference can be prevented by enclosing the triac and control components in a metal case, and interference carried through the leads may be filtered out by chokes (inductances) in series, and capacitors in parallel.

EXPERIMENT 11.3: Semiconductor infra-red source and detector, modulated infra-red, and "light guide" ("Crofon").

Apparatus required:

Gallium arsenide (GaAs) light source type MGA100 (Proops Bros.).
Infra-red detector type MSP3 (Proops).
Ammeter, f.s.d. 1 A.

AF oscillator with low-impedance (e.g. 3 Ω) output.
3 V 1 amp D.C. supply.
20 V 50 mA D.C. supply (separate).
Milliammeter, f.s.d. 10 mA.
A.F. amplifier with loudspeaker.
Short length of flexible light guard (Proops).

Procedure. Set up the arrangement of Figure 11.29. At first place the GaAs source very close to the detector; check that the source, which must be in good thermal contact with a large metal "heat-sink", is not passing more than about half an ampere. If it gets too hot to touch it may be irreparably damaged.

The receiver meter should read about 1 mA, and the oscillator note should be heard in the loudspeaker.

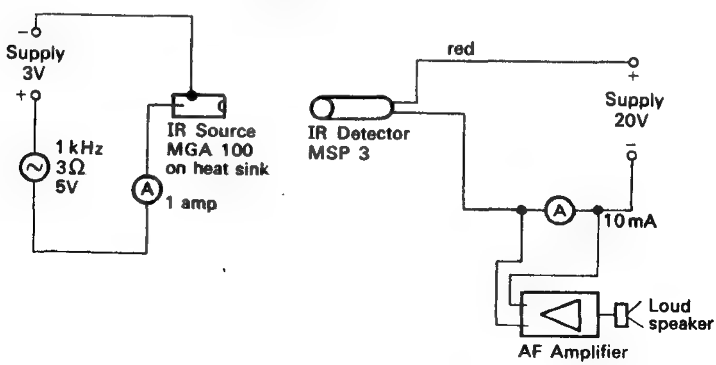


FIG. 11.29.

Now separate the source and detector, and apply opposite ends of about 1 foot of light-guide to the openings of each. Although much attenuated, the signal should again be heard, showing that the invisible infra-red beam is carrying the audio-frequency signal, and that the beam can be carried round corners by the light-guide.

EXPERIMENT 11.4: The phototransistor.

The "Unilab" mounting of the OCP71 is very convenient. Set up the circuit of Figure 11.30 and observe the effect of bringing a 12-V 6-W bulb up to the phototransistor, from the direction indicated by an arrow.

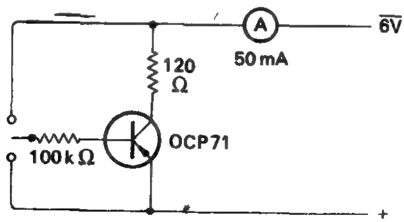


FIG. 11.30.

The phototransistor base has so far been left "floating". It is more often connected to the emitter by a 100 kΩ resistor. Study the effect of connecting it first to the negative supply line, then to the emitter, through a 100 kΩ resistor.

Further projects. Investigate the Intensity of Illumination Collector Current Graphs, as in Expt. 4.1; try both of the above base connections. Which gives the most convenient light-meter characteristic?

EXPERIMENT 11.5: A unijunction relaxation oscillator.

Apparatus required:

Unijunction transistor, e.g. Texas Instruments type 2N1671.
Resistors, 100 kΩ and 30 Ω.
Capacitor, 50 F and 1 F.
Supply, 15 V D.C.
Oscilloscope.

Procedure. Set up the circuit of Figure 11.31. Connect the oscilloscope input in turn to points *A* and *B*, where rectangular and triangular waveforms will be found.

Alteration of the 100 kΩ resistor and/or the 50 μF capacitor will alter the fundamental frequency.

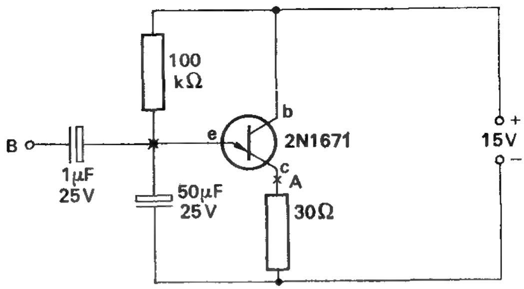


FIG. 11.31.

Try also the very similar metronome circuit of Figure 11.17.

EXPERIMENT 11.6: Drain characteristics of an FET.

Apparatus required:

Low-power *n*-channel FET, e.g. Texas Instruments TIS14.
D.C. supplies, 6 V and 20 V.
Voltmeters, f.s.d. 6 V and 20 V.
Milliammeter, f.s.d. 20 mA.
Potentiometers, 1000 Ω and 100 Ω.

Procedure. Set up the circuit of Figure 11.32 (compare with Vol. 1 p. 123: anode characteristics of a thermionic triode). For various fixed values of V_{GS} plot graphs of drain current I_D against source-drain voltage V_{DS}.

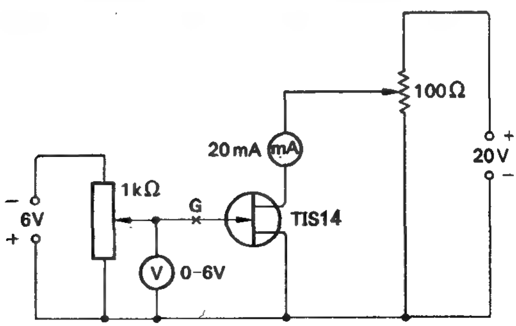


FIG. 11.32.

Compare the set of curves so obtained with those of a thermionic pentode (Vol. I, pp. 133 and 138).

From the slopes of the linear portions of the graphs, obtain values for the output impedance of the FET $\left(\frac{\text{change in drain voltage}}{\text{change in drain current}}\right)$.

Further investigation. A galvanometer may be inserted at point *G*, to determine the gate current at various gate and drain voltages. The input resistance, $\frac{\text{change in gate voltage}}{\text{change in gate current}}$, may then be determined; the main advantag of the FET lies in the fact that this figure is very high, typically of the order of 5 MΩ.

EXPERIMENT 11.6.2: Photovoltaic effect.

Apparatus required:

Solar cell, e.g. International Rectifier type S4M.
Voltmeter, 0–5 V; Ammeter, 100 mA or 1 A.

Connect the cell direct first to the ammeter, then to a voltmeter. Place it close to a 60-W lamp, and note that it can supply nearly 1 V, or a current of about 100 mA. (The S4M contains four cells in parallel; other types have greater or lesser outputs.)

EXPERIMENT 11.7. Function of a typical integrated circuit.

Apparatus needed:

Integrated Circuit (RCA) type CA3020 (Proops Ltd.).
Piece of "Veroboard", 12×9 matrix (nine copper strips with twelve holes on each).
2 resistors, 68 Ω.
1 capacitor, 10 μF 12 V.
1 capacitor, 1 μF 12 V.
Supply, 6 V D.C. 50 mA.
A.F. oscillator.
Oscilloscope or valve voltmeter.

Procedure. Connect the integrated circuit using the Veroboard as shown in Figure 11.33, or similarly. The aim is not smallness, but easily followed circuitry.

Connect the supply as shown, and from the A.F. oscillator apply about 5 mV between input and supply negative. Measure or 'scope across the output leads. Calculate the voltage and power gain, and repeat at various frequencies. If suitable test equipment is available, the gain should be found to be still substantially constant at up to 10 MHz.

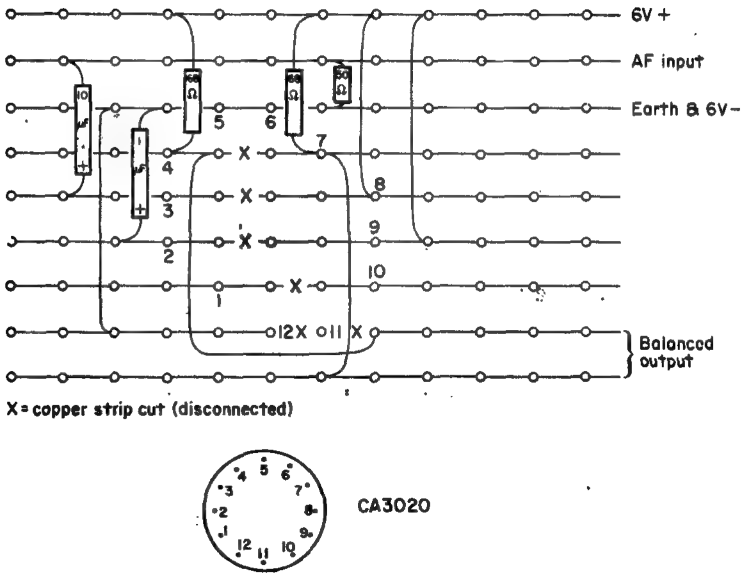


FIG. 11.33.

QUESTIONS

1. What features of a heterojunction make it (a) little use as a rectifier and (b) useful as a very high-speed switch?

2. A small D.C. motor is controlled by being connected in series with a D.C. supply of 25 V, and the collector and emitter leads of a power *p–n–p* transistor whose current amplification is 30. (a) Find the power dissipated as heat in the transistor when the motor current is 1 A and the voltage across the motor is 18 V; (b) find the approximate resistance which would have to be connected between supply negative and base to produce this motor current.

3. A thyristor is connected in series with a lamp whose brightness is to be controlled, and the A.C. mains. (a) Sketch roughly the lamp current waveform when the lamp is dissipating about one-quarter maximum power; (b) say why the lamp flickers more noticeably when on low brightness; (c) compare the average lamp power when the thyristor is short-circuited with that when the thyristor control is in circuit and full on.

4. A certain tunnel diode, suitably biased, shows a negative resistance equivalent to $-30\ \Omega$. What is the effective resistance when this is connected (a) in series with a $25\ \Omega$ resistance, (b) in parallel with a $33\ \Omega$ resistance?

5. Calculate the efficiency as a light source of a gallium arsenide wafer 1 mm square, forward biased at 3·3 V and carrying a current of 500 mA, and emitting 750 μW of electromagnetic radiation.

What practical steps would have to be taken to prevent the device suffering damage in these circumstances? Suggest suitable figures.

6. Compare the advantages of (a) a cadmium sulphide photoconductive cell, (b) a photovoltaic semiconductor device, (c) a phototransistor, for (i) light measurement for photography, (ii) location or counting of small objects, (iii) use in a burglar-alarm system.

7. What are the advantages of field-effect transistors over other types for use in instruments for measuring very small currents (less than 10^{-10} A)? What special precautions have to be taken when installing or using FETs?

8. Explain what is meant by the following terms: (i) zone refining, (ii) alloy junction, (iii) epitaxial growth, (iv) planar construction. Explain briefly how use is made of some of these techniques, with others, to produce a microcircuit.

(U.L. B.Sc. (Eng.) Part 2, Electronics, 1967)

APPENDIX I

THE h-PARAMETERS

THE h or hybrid parameters are relationships between the input and output currents and voltages of any four-terminal network (e.g. a transistor or a valve). In Figure I.1, the four-terminal network is shown as a "black box" with a voltage v_1 and current i_1 applied at the input, and a voltage v_2 and current i_2 at the output. The directions of the currents are as indicated; the polarities of v_1 and v_2 are such that terminal 1 is positive with respect to terminal 2, and terminal 3 is + w.r.t. terminal 4. v and i stand for the *alternating components* of the voltages and currents concerned; any steady values present are ignored.

i_1 and v_2 are taken as the independent variables, and we define them by the equations

$$v_1 = h_{11}i_1 + h_{12}v_2 \tag{I.1}$$

$$i_2 = h_{21}i_1 + h_{22}v_2 \tag{I.2}$$

It is then clear from these equations that eqn. (I.1) represents the input side of the black box, and eqn.(I.2) the output. Figure I.2 shows the electrical

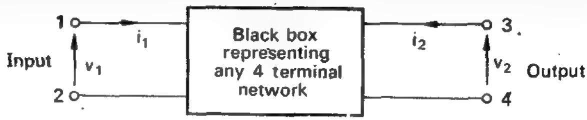


FIG. I.1.

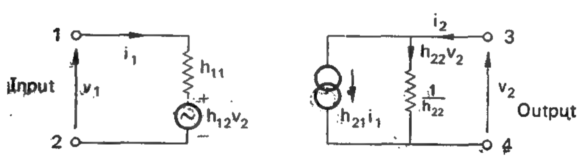


FIG. I.2.

equivalent circuit of the black box in terms of these h-parameters. The input circuit is a resistance of h_{11} in series with a voltage generator of magnitude $h_{12}v_2$, and is thus dependent on the output circuit. The output circuit is an admittance of h_{22} (or resistance of h_{22}^{-1}) in parallel with a current generator of magnitude $h_{21}i_1$, and is thus dependent on the input circuit.

From eqn. (I.1),

$$h_{11} = \frac{v_1}{i_1} \quad \text{when} \quad v_2 = 0.$$

This means that h_{11} is the ratio of the change in the steady value of v_1 to the corresponding change in the steady value of i_1 when the value of v_2 is kept constant.

So, $$h_{11} = \left(\frac{\delta V_1}{\delta I_1}\right)_{V_2} = \text{input impedance when the output is short-circuited to A.C}$$

Similarly, $$h_{12} = \left(\frac{\delta V_1}{\delta V_2}\right)_{I_1} = \text{reverse voltage feedback ratio when the input is open-circuited to A.C.}$$

$$h_{21} = \left(\frac{\delta I_2}{\delta I_1}\right)_{V_2} = \text{forward current gain with output short-circuited to A.C.}$$

$$h_{22} = \left(\frac{\delta I_2}{\delta V_2}\right)_{I_1} = \text{output admittance when the input is open-circuited to A.C.}$$

The h-parameters for a transistor in common emitter configuration

The input impedance parameter h_{11} when applied to a transistor with grounded emitter is

$$h'_{11} = \left(\frac{\delta V_{EB}}{\delta I_B}\right)_{V_{CE}} = h_{ie}$$

where the prime above the h_{11} is the notation adopted to indicate that the emitter is grounded. (If the base is grounded, there is no dash; two dashes means grounded collector.) An alternative method is to use the symbol h_{ie}, where the i refers to input, and the e to the grounded terminal. Thus, h'_{11} means the ratio of change in emitter–base voltage to the corresponding change in base current, when the collector–emitter voltage is kept constant. Similarly,

$$h'_{12} = \left(\frac{\delta V_{EB}}{\delta V_{CE}}\right)_{I_B} = h_{re}$$

where the r in h_{re} refers to "reverse";

$$h'_{21} = \left(\frac{\delta I_c}{\delta I_B}\right)_{V_{CE}} = h_{fe}$$

where the f in h_{fe} refers to "*f*orward", and

$$h'_{22} = \left(\frac{\delta I_c}{\delta V_{CE}}\right)_{I_B} = h_{oe}$$

where the o in h_{oe} refers to "*o*utput".

The values of these parameters may be determined to better than 1% for an individual transistor, but may vary by a factor of 4 between transistors of the same type-number. Consequently, accurate calculations from available characteristics are pointless, and frequent approximations are justified. It is important to realize that, even for a *given* transistor, the *h*-parameters are not constants; they are quoted at stated values of V_{CE}, I_B and frequency, and are different at different such values.

APPENDIX II

RELATIONSHIP BETWEEN *h* AND *T* PARAMETERS FOR COMMON EMITTER CONFIGURATION

In Figure II.1 we have redrawn Figure 9.2(a) with the directions of both i_B and i_C reversed so that we can draw a direct comparison with Figure II.1. The latter is just Figure I.2 redrawn with specific reference to a transistor operating with grounded emitter.

For the equivalent T circuit in Figure II.1:

$$v_{BE} = i_B r_b - i_E r_e \qquad \text{But} \qquad i_E = -(i_B + i_c)$$

$$\therefore\ v_{BE} = i_B r_b + (i_B + i_c) r_e = i_B(r_e + r_b) + i_c r_e \tag{II.1}$$

and

$$\begin{aligned} v_{CE} &= i_c r_c + \alpha i_E r_c - i_E r_e = i_c r_c + \alpha i_E r_c + (i_B + i_c) r_e \\ &= i_c r_c + \alpha r_c(-i_B - i\ ) + (i_B + i_c) r_e \\ &= i_B(r_e - \alpha r_c) + i_c[r_e + r_c(1-\alpha)] \end{aligned} \tag{II.2}$$

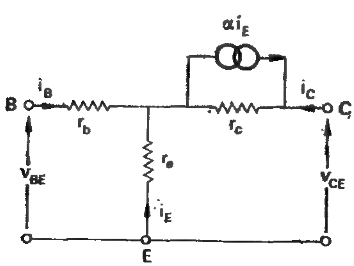


Fig. II.1.

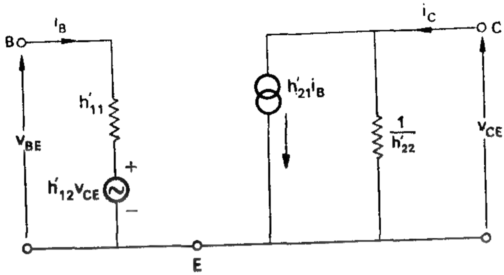


FIG. II.2.

For the equivalent h circuit in Figure II.2:

$$v_{BE} = i_B h'_{11} + h'_{12} v_{CE} \tag{II.3}$$

and

$$i_c = i_B h'_{21} + h'_{22} v_{CE}$$

From the last equation,

$$v_{CE} = \frac{i_c}{h'_{22}} - i_B \cdot \frac{h'_{21}}{h'_{22}}$$

$$= i_B \left(-\frac{h'_{21}}{h'_{22}}\right) + i_c \left(\frac{1}{h'_{22}}\right) \tag{II.4}$$

Substituting for V_{CE} in (II.3) we have:

$$v_{BE} = i_B h'_{11} + h'_{12}\left[-\frac{h'_{21}}{h'_{22}}\, i_B + \frac{i_c}{h'_{22}}\right]$$

$$= i_B \left[h'_{11} - \frac{h'_{12}h'_{21}}{h'_{22}}\right] + i_c \frac{h'_{12}}{h'_{22}} \tag{II.5}$$

By equating coefficients of i_B in eqns. (II.2) and (II.4) we obtain:

$$r_e - \alpha r_c = -\frac{h'_{21}}{h'_{22}} \tag{II.6}$$

Similarly for i_C:

$$r_e + r_c(1-\alpha) = \frac{1}{h'_{22}} \tag{II.7}$$

By equating coefficients of i_B in eqns. (II.1) and (II.5) we have:

$$r_e + r_b = h'_{11} - \frac{h'_{12}h'_{21}}{h'_{22}} \tag{II.8}$$

Similarly for i_C:

$$r_e = \frac{h'_{12}}{h'_{22}} \tag{II.9}$$

Substituting for r_e from (II.9) into (II.8):

$$r_b = h'_{11} - \frac{h'_{12}h'_{21}}{h'_{22}} - \frac{h'_{12}}{h'_{22}}$$

$$= h'_{11} - \frac{h'_{12}(1+h'_{21})}{h'_{22}}$$

Substituting for r_e from (II.9) into (II.6):

$$\frac{h'_{12}}{h'_{22}} - \alpha r_c = \frac{-h'_{21}}{h'_{22}} \quad \therefore \alpha = \left(\frac{h'_{12}}{h'_{22}} + \frac{h'_{21}}{h'_{22}}\right)\frac{1}{r_c}$$

Substituting for α and r_e into (II.7):

$$\frac{h'_{12}}{h'_{22}} + r_c\left[1 - \frac{1}{r_c}\left(\frac{h'_{12}}{h'_{22}} + \frac{h'_{21}}{h'_{22}}\right)\right] = \frac{1}{h'_{22}}$$

$$\therefore \frac{h'_{12}}{h'_{22}} + r_c - \left(\frac{h'_{12}}{h'_{22}} + \frac{h'_{21}}{h'_{22}}\right) = \frac{1}{h'_{22}}$$

$$\therefore r_c = \frac{1}{h'_{22}} - \frac{h'_{12}}{h'_{22}} + \frac{h'_{12}}{h'_{22}} + \frac{h'_{21}}{h'_{22}}$$

$$\therefore r_c = \frac{1+h'_{21}}{h'_{22}}$$

Substituting back for α:

$$\alpha = \left(\frac{h'_{12}}{h'_{22}} + \frac{h'_{21}}{h'_{22}}\right)\frac{h'_{22}}{1+h'_{21}}$$

$$\therefore \alpha = \frac{h'_{21}+h'_{12}}{1+h'_{21}}$$

Summary

$$r_e = \frac{h'_{12}}{h'_{22}} = \frac{h_{re}}{h_{oe}}.$$

$$r_b = h'_{11} - \frac{h'_{12}(1+h'_{21})}{h'_{22}}$$

$$= h_{ie} - \frac{h_{re}(1+h_{fe})}{h_{oe}}$$

$$r_c = \frac{1+h'_{21}}{h'_{22}} = \frac{1+h_{fe}}{h_{oe}}$$

$$\alpha = \frac{h'_{21}+h'_{12}}{1+h'_{21}} = \frac{h_{fe}+h_{re}}{1+h_{fe}}$$

It is these conversions which are used in Section 9.5.

APPENDIX III

INPUT AND OUTPUT RESISTANCES AND GAINS FOR GROUNDED EMITTER TRANSISTOR, IN TERMS OF *h*-PARAMETERS

FIGURE III.1 shows an alternating voltage source V_S in series with its internal resistance R_S applied to the input terminals of a two-port network (represented by a "black box") and across the output terminals is a load resistance of $G_L^{-1}\Omega$, where G_L is the admittance of the load in mhos. Then we have:

For the input circuit, $$v_S - i_1 R_S = v_1 \tag{III.1}$$

For the output circuit, $$i_2 = -v_2 G_L \tag{III.2}$$

Also, we have: $$v_1 = h'_{11} i_1 + h'_{12} v_2 \tag{III.3}$$

and $$i_2 = h'_{21} i_1 + h'_{22} v_2 \tag{III.4}$$

From (III.4), $$v_2 = \frac{i_2 - h'_{21} i_1}{h'_{22}}$$

From (III.2), $$v_2 = \frac{-i_2}{G_L}$$

$$\therefore \frac{-i_2}{G_L} = \frac{i_2 - h'_{21} i_1}{h'_{22}}$$

$$\therefore \text{ current gain } A_I = \frac{i_2}{i_1} = \frac{h'_{21} G_L}{h'_{22} + G_L} \tag{III.5}$$

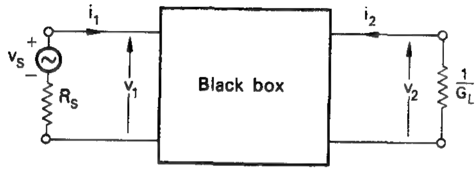


FIG. III.1.

Substituting for i_2 from (III.5) into (III.2),

$$\frac{h'_{21}G_L i_1}{h'_{22}+G_L} = -v_2 G_L$$

so

$$v_2 = \frac{-h'_{21}i_1}{h'_{22}+G_L} \tag{III.6}$$

Substituting for v_2 from (III.6) into (III.3),

$$v_1 = h'_{11}i_1 + h'_{12}\left(\frac{-h'_{21}i_1}{h'_{22}+G_L}\right)$$

$\therefore$ input resistance

$$R_{\text{in}} = \frac{v_1}{i_1} = h'_{11} - \frac{h'_{12}h'_{21}}{h'_{22}+G_L} \tag{III.7}$$

The voltage gain

$$A_V = \frac{v_2}{v_1} = \frac{-i_2}{v_1 G_L}$$

Substituting for v_1 from (III.3),

$$A_V = \frac{-i_2/G_L}{h'_{11}i_1 + h'_{12}v_2} = \frac{-i_2/G_L}{h'_{11}i_1 + h'_{12}(-i_2/G_L)}$$

Substituting for i_1 from (III.5),

$$A_V = \frac{-i_2/G_L}{h'_{11}i_2(h'_{22}+G_L)/(h'_{21}G_L) - h'_{12}(i_2)/(G_L)}$$

$$= \frac{-1}{(h'_{11}/h'_{21})(h'_{22}+G_L) - h'_{12}}$$

$$= \frac{-h'_{21}}{h'_{11}G_L + h'_{11}h'_{22} - h'_{12}h'_{21}}$$

Let

$$\Delta = h'_{11}h'_{22} - h'_{12}h'_{21},$$

then voltage gain

$$= A_V = \frac{-h'_{21}}{h'_{11}G_L + \Delta} \tag{III.8}$$

The power gain A_P is just the product of the current and voltage gains:

$$A_P = A_I A_V$$
$$= \frac{(h'_{21})^2 G_L}{(h'_{22}+G_L)(h'_{11}G_L+\Delta)} \tag{III.9}$$

The output resistance R_{out} is found with least trouble by considering a very large value of load resistance, such that G_L tends to zero. Ideally this is an open-circuit output, such that the output voltage is constant. Under this condition, by definition,

$$R_{in} = \frac{v_1}{i_1} = h'_{11} \tag{III.10}$$

and

$$i_2/i_1 = h'_{21} \tag{III.11}$$

But, for the input circuit, $v_1 = i_1 R_{in}$.

Substituting for v_1 in (III.1),

$$v_S - i_1 R_S = i_1 R_{in}$$
$$\therefore\ i_1 = \frac{v_S}{R_S+R_{in}} \tag{III.12}$$

Substituting for i_1 in (III.11),

$$i_2 = \frac{h'_{21} v_S}{R_S+R_{in}}$$

Substituting for R_{in} from (III.10):

$$i_2 = \frac{h'_{21} v_S}{R_S+h'_{11}} \tag{III.13}$$

From (III.4),

$$v_2 = \frac{i_2 - h'_{21} i_1}{h'_{22}}$$

Substituting for i_2 from (III.5), and for i_1 from (III.12):

$$v_2 = \frac{[(h'_{21}G_L)/(h'_{22}+G_L) - h'_{21}]\, v_S/(R_S+R_{in})}{h'_{22}}$$
$$= \frac{(h'_{21}G_L - h'_{21}h'_{22} - h'_{21}G_L)\, v_S}{h'_{22}(h'_{22}+G_L)((R_S+h'_{11}) - (h'_{12}h'_{21})/(h'_{22}+G_L))}$$

where we have used the value of R_{in} from (III.7).

As $\quad G_L = 0, \quad v_2 = \dfrac{-h'_{21} v_S}{h'_{22}((R_S+h'_{11}) - (h'_{12}h'_{21})/(h'_{22}))} \quad$ (III.14)

The output resistance R_{out} is found by dividing v_2 from (III.14) by i_2 from (III.13):

$$R_{out} = \frac{-v_2}{i_2} = \frac{R_S+h'_{11}}{R_S h'_{22}+\Delta} \tag{III.15}$$

It is interesting to substitute the data of the JJ1 transistor of Section 9.5, and check the values. We used

$$h'_{11} = 3500\ \Omega, \qquad h'_{12} = 25\times10^{-4},$$
$$h'_{21} = 49, \qquad h'_{22} = 50\times10^{-6}\ \text{mho};$$
$$G_L = 10^{-4}\ \text{mho}; \qquad R_S = 10^3\ \Omega.$$

Substituting in (III.5),

$$\text{Current gain } A_I = \frac{h'_{21}G_L}{h'_{22}+G_L}$$
$$= \frac{49\times10^{-4}}{50\times10^{-6}+10^{-4}} = \frac{49\times10^{-4}}{150\times10^{-6}} = 32{\cdot}7$$

Substituting in (III.7),

$$\text{Input resistance} \qquad R_{\text{in}} = h'_{11} - \frac{h'_{12}h'_{21}}{h'_{22}+G_L}$$
$$= 3500 - \frac{25\times10^{-4}\times49}{50\times10^{-6}+10^{-4}} = 3500 - \frac{250\times49}{15} = 3500-820 = \mathbf{2680}\ \Omega.$$

Substituting in (III.8),

$$\text{Voltage gain} = \frac{-h'_{21}}{h'_{11}G_L+\Delta} \quad \text{where} \quad \Delta = h'_{11}h'_{22} - h'_{12}h'_{21}$$
$$\Delta = 3500\times50\times10^{-6} - 25\times10^{-4}\times49 = 0{\cdot}1750 - 0{\cdot}1225 = 0{\cdot}0525$$
$$\therefore\ A_V = \frac{-49}{3500\times10^{-4}+0{\cdot}0525} = \frac{-49}{0{\cdot}4025} = -124$$

Substituting in (III.15):

$$R_{\text{out}} = \frac{R_S+h'_{11}}{R_S h'_{22}+\Delta} = \frac{10^3+3500}{10^3\times50\times10^{-6}+0{\cdot}0525} = \frac{4500}{0{\cdot}1025} = 44\ \text{k}\Omega.$$

Note that, as expected, these values are precisely the same as those found by equivalent T parameters in Section 9.5.

APPENDIX IV

CLASSIFICATION OF SEMICONDUCTOR TYPES

OLD METHOD

The *first* letter indicated the device; 0 meant that it was some kind o semiconductor. (Other letters were used for valves, e.g. EL84, UU5.)

The *second* letter was A for a diode, or C for a transistor, e.g.

OA81 is a semiconductor diode;
OC83 is a transistor.

NEW METHOD

As research and development took place the old system was found inadequate. Under the new system, the *first* letter is A for a germanium semiconductor, B for a silicon one, and R for a compound material.

The significance of the *second* letter is:

A: diode,
C: low-power A.F. transistor,
D: power A.F.,
F: R.F.,
L: R.F. power,
P: photosensitive,
R: controlling and switching device having a specified breakdown characteristic,

S: switching transistor,
T: a power version of R,
Y: a power rectifier diode,
Z: a Zener or reference diode,
e.g. AF114 is a germanium R.F. transistor;
BYY15 is a silicon rectifier.

APPENDIX V

PRACTICAL PRECAUTIONS WITH TRANSISTORS

1. *Soldering*

Poor soldering techniques are more likely to result in damage when semiconductor devices are included in the circuit, so it is necessary in the first place to make sure you are soldering in the correct way.

Generally, two or more leads and/or tags are to be soldered together. All pieces to be joined must first be heated above the melting-point of the solder, by pressing them all together and against the bit of the iron. Then, keeping the iron in contact, melt new solder onto the joint. Quickly remove the iron, and see that the pieces of the joint do not move until the solder has solidified. The solder must melt *at* the joint, in contact with all parts to be joined, and all parts must be above the melting point. The usual mistake is to carry already molten solder on the iron to a joint which is not hot enough. Such a joint can look and feel satisfactory, but have a high or intermittent electrical resistance; it is called a "dry joint", and is one of the most difficult faults to locate.

2. *Keeping the component cool*

Semiconductors, and some electrolytic capacitors, are irreparably damaged by too much heat, so the soldering process must be achieved quickly and at the first attempt. To minimize the heat carried along a transistor lead to the junctions, use a "thermal shunt". This consists of a large metal item firmly clamped to the lead, between the component and the point of soldering. Convenient shunts are a small pair of pliers, a crocodile clip, a paper clip, or one of the heat clips specially made for the purpose.

If transistor leads are cut short, as is often desirable, heat shunting becomes more important.

An unsatisfactory soldering attempt must be allowed to cool before another attempt is made.

3. *Fets, Igfets and Mosfets*

Field effect transistors often incorporate an insulating layer consisting of only a few molecules thickness. Such layers can be destroyed by very small induced e.m.f.s or currents. For example, FETs are often broken merely by having their leads touched. The e.m.f. induced in the human body by the A.C. mains field is sufficient to do the damage, so are "contact" and "thermal" potentials. So a FET must be handled with all its leads shorted close to the body, by a clip of copper wire which is only removed after the FET has been completely wired into the circuit.

4. *Quick-heat soldering irons*

In these, the "bit" is a thick wire which carries several amps from the secondary of a small transformer in the body. Of course, such a large current can induce in nearby circuits more than enough current to wreck a FET, in fact often sufficient to damage a robust conventional transistor. The safest rule is not to use such an iron on circuits containing small transistors.

5. *Unearthed soldering irons*

A small leakage current will flow to earth through any circuit touched by a mains-operated soldering iron whose metal parts are not earthed. Such currents have often resulted in transistor damage. Soldering irons must be earthed, and properly connected to a three-pin supply: but remember not to use them on circuits which are live!

6. *Heat sinks*

Nearly all transistors in power circuits require a heat sink to dissipate the electrical energy which turns to heat at the junctions. Heat sinks range from a couple of square centimetres of thin copper plate, to large, finned metal blocks.

The effectiveness of a heat sink is measured in watts/°C; for example, a

small transistor with only its own case to dissipate the heat may have a thermal conductance of 0·4 mW/°C, whereas a large sink might be rated at 3 W/°C.

Taking, for example, a system of 2 W/°C, if 2 W are being dissipated it will run at a temperature of only 1°C above room (ambient) temperature; or if 20 W are being released in the transistor, it will reach 10°C above room temperature. Conversely, if room temperature is 20°C, and the specified maximum working temperature of the device is 70°C, the temperature excess can be allowed to rise to 50°C, which will occur if 100 W are being dissipated.

Alternatively, the manufacturer may quote a "derating". For example, he may (for the same device as above) say that the dissipation at 20°C may be 100 W; derate linearly to 0 W at 70°C. (Or, derate linearly at 2 W/°C.) From this we can see that if the ambient temperature is 30°C, an internal dissipation of 80 W could be allowed.

In audio-amplifiers having outputs of 5 W or over, it is often found adequate to arrange for the dissipation of the power transistors to be spread through the metal framework of the amplifier. Usually this entails bolting the transistor to the frame in such a way that good thermal contact is made, at the same time as good electrical insulation. Washers of mica, or thin polythene films, are used for this purpose.

7. *Maximum current ratings*

Failure of some other component in an amplifier circuit could easily result in the transistor becoming "bottomed", or turned full on. Typically, the electrolytic connecting the base of a common emitter amplifier might go "short-circuit", applying a high forward voltage to the base.

Damage to the transistor in these circumstances will be prevented if the total resistance in emitter and collector leads prevents the maximum allowed current from being exceeded. For example, the maximum collector current for the ACY17 is $\frac{1}{2}$ A. If a 50-V supply is in use, a total resistance in emitter and collector leads of 100 Ω will ensure that this current is not exceeded, even if something goes wrong elsewhere in the circuit.

Some such resistance should be included in all experimental circuits.

APPENDIX VI

SYMBOLS USED IN THE TEXT

A_I	Current amplification	—
A_P	Power amplification	—
A_V	Voltage amplification	—
B	Magnetic flux density	Wb m^{-2}
E	Electric field intensity	V m^{-1}
h_{11}, h_{ce}, h'_{21}, etc.	See Appendix I	
I_B	D.C. value of base current	A
I_b	A.C. value of base current	A
I_C, I_c	D.C. and A.C. value of collector current	A
I_E, I_e	D.C. and A.C. value of emitter current	A
I_G, I_g	D.C. and A.C. value of gate (or trigger) current	A
I_S, I_s	D.C. and A.C. value of source current	A
I_{CBO}	Collector leakage current when collector is reverse-biased w.r.t. base, and emitter is open-circuited	A
I_{CEO}	Collector leakage current when collector is reverse-biased w.r.t. emitter, and base is open-circuited	A
I	Current	A
J	Current density	A m^{-2}
k	Boltzmann's constant $1{\cdot}38 \times 10^{-23}$	J K^{-1}
L_e	Diffusion length for electrons into emitter	m
L_h	Diffusion length for holes into base	m
m	Mass	kg
n	Free electron carrier density (also any pure number)	m^{-3}
p	Hole carrier density	m^{-3}
P	Pressure	N m^{-2}
q	Charge	C

Q	Q-factor, quality factor	—
R_{in}	Input resistance	Ω
R_L	Load resistance	Ω
R_{out}	Output resistance	Ω
r_b	Base resistance	Ω
r_c	Collector resistance	Ω
r_e	Emitter resistance	Ω
S	Stability factor	—
t	Time, thickness	s, m
T	Absolute temperature	°K
u	Velocity	m s^{-1}
V	Voltage, p.d., e.m.f.	V
V_{CE}	D.C. voltage between collector and emitter; C positive	V
V_{ce}	A.C. voltage between collector and emitter	V
V_{BE}, V_{be}, etc.	likewise	V
V_{DS}, V_{GS}	D.C. values of drain-source voltage, gate-source voltage	V
V_{in}	A.C. value of input voltage	V
W_g	Ionization energy	eV
W	Width	m
α	Current gain in common base configuration	—
β	Current gain in common emitter configuration	—
β	Transport factor	—
γ	Emitter efficiency	—
δ	Collector efficiency	—
Δx	Change in the value of variable x	—
λ	Mean free path	m
μ	Mobility	$m^2 V^{-1} sec^{-1}$
ϱ	Resistivity	Ω m
σ	Electrical conductivity	$\Omega^{-1} m^{-1}$

Abbreviation of Units

A	Ampere
C	Coulomb
eV	Electronvolt
kg	Kilogramme
m	Metre
N	Newton
s	Second
V	Volt

APPENDIX VI

J		Joule
°K		Degree Kelvin
Wb		Weber
Ω		Ohm
GHz	=	10^9 Hz

APPENDIX VII

IMPURITY SEMICONDUCTORS

THE two common semiconductor materials are germanium and silicon: these are both Group IV elements, which means that each atom has four valence electrons and can make four bonds with neighbouring atoms. A perfect crystal of (say) silicon is shown in Figure VII.1.

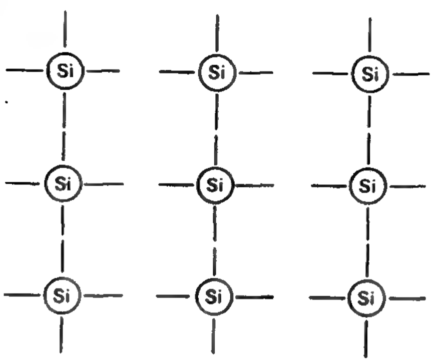


FIG. VII.1. A perfect crystal lattice of silicon atoms

If perhaps one in a million of these silicon atoms is replaced by an impurity such as phosphorus, which has a valency of 5, then the final crystal lattice structure has an extra electron near the site of the phosphorus atom. The electrical neutrality of the original silicon has now been lost, as it is now slightly negative (see Fig. VII.2). The semiconductor is now *n*-type, and the phosphorus atom is a DO*N*OR impurity because it has donated

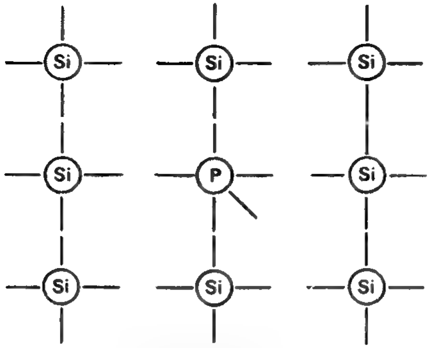


FIG. VII.2. Some n-type silicon (due to the addition of phosphorus, a donor impurity)

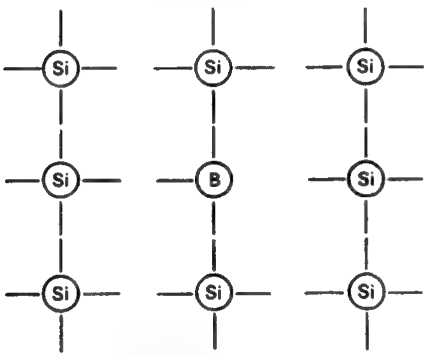


FIG. VII.3. Some p-type silicon (due to the addition of boron, an acceptor impurity)

an electron to the lattice. If, however, a boron atom, of valency 3, replaced the phosphorus atom, then the semiconductor would be slightly positive (p-type) and the boron atom is called an ACCE*P*TOR impurity because it can accept an electron (see Fig. VII.3). Note that if boron can accept an electron into its valency structure it is behaving as if a hole is present and this hole has a positive property (because it is the ***absence*** of a] ***negative*** charge).

BIBLIOGRAPHY

JAMES J. BROPHY, *Semiconductor Devices.* George Allen and Unwin.
A. J. DEKKER, *Electrical Engineering Materials.* Prentice–Hall.
DEWITT and ROSSOFF, *Transistor Electronics*, McGraw–Hill.
R. G. HIBBERD, *Transistor Pocket Book.* Newnes.
ALAN HOLDEN, *Conductors and Semiconductors.* Bell Telephone Laboratories.
JENKINS and JARVIS, *Basic Principles of Electronics,* Volume I. Pergamon.
Reference Manual of Transistor Circuits, pub. Mullard Ltd.

INDEX TO SECTIONS

CONTENTS OF THIS BOOK

08 016118 9

JENKINS/JARVIS

BASIC PRINCIPLES OF ELECTRONICS
VOLUME 2 SEMICONDUCTORS